

What will emerge from Tokyo?

The economic summit meeting that opens in Tokyo on 4 May could be a chance to put the West's affairs in better order. But the precedents are discouraging.

THERE could be no better place than Tokyo for this year's meeting of the leaders of the seven large industrialized nations of the West. Although the city may not yet be the financial capital of the world, Japan as a whole has become the world's industrial pacemaker. Even as the participants are shuttled quickly from one event to another, they will not be able to avoid the question of how the most orderly of all the world's metropolitan conurbations can also be the source of repeated innovation which, elsewhere, is more usually marked by chaos. Those who can afford the time should stay on for a day or so after the two-day meeting; they will not be able in that time to come to grips with all the secrets of Japan's success, but they might be able to pick up a few tips that could help with the domestic problems awaiting their return and which will be waiting unchanged, as they have been for the past decade, unless this year's summit is markedly more successful than its predecessors.

Tokyo is planned as a largely economic meeting, and for good reason. These meetings became regular and annual only after the price of oil shot up in 1973; they were well established in the top politicians' calendar by the time of the 1979 meeting in Venice, when those present could do little but wring their hands about the second round of the price increase decreed by the Organisation of Petroleum Exporting Countries (OPEC). By that token, this year's meeting could have been an occasion for celebrating the demise of OPEC, at least temporarily, and for planning what should happen next. If the United States can avoid reading too many lessons to too many people about the horrors of terrorism, that could be a useful exercise.

Balance

What the Tokyo meeting might usefully do is to strike a better balance between the optimists and the pessimists about the likely course of events. The collapse of OPEC's policies has wonderfully increased the population of the optimists, who are especially thick on the ground where there are equity shares in industrial companies to buy and sell, and whose eagerness to buy has driven most of the world's stock exchanges to record price peaks (which means that shareholders forgo income now in the belief that there are capital gains to come). Nobody can conclusively quarrel with the view that the world stands on the threshold of a new period of economic growth from which everybody will benefit. The snag is that there is no evidence conclusively to support this cheerful view, nor sufficient evidence for the opposite conclusion that this may be the threshold of a time of honours. For the pessimists' case is not so much a case as a collection of gloomy speculations: what if countries such as Mexico should declare themselves bankrupt because selling oil at reduced prices will not allow them to service their debt? What if the US dollar, which now buys only two-thirds as many Japanese yen as at the beginning of 1985, should keep on falling? Or if the industrialized countries fearful of losing their place in the competition should seek to save themselves by trade protection, beggaring everybody in the process? A few years ago, when catastrophe theory in mathematics was fashionable, this would be recognized as a time of incipient bifurcation: events could go one way or the other.

World leaders may be no better able than the rest of us to tell which view is right, but what they decide in Tokyo could help to show which way the dice will fall. There are three groups of issues to be settled: money, trade and debt. The largest countries have been congratulating themselves in the past few months for having using their inner club (called the group of five) successfully to make the value of the US dollar decline. Central banks seem to have been able to trigger a substantial movement at very little cost, but may in the process have destroyed the source of the dollar's previous strength, the belief from which only a few sceptics dissented that it would always be a rock-solid currency. The problem now is whether bankers could as cheaply stop the continuing slide if they thought that necessary.

Usefulness

The devaluation will serve three useful purposes. By making imported goods and services more expensive in the United States, it will reduce the external deficit of the United States. Indirectly, it may help to simplify the funding of the US budget deficit, which means that interest rates will be given a further downward push. And it will mark a step towards realism, the recognition that industrial power has partly migrated from the United States, almost certainly for good. Even for the United States, this is a step towards enlightenment. The grounds for pessimism are different, and concern the capacity of the world's institutions, mostly financial institutions, to accommodate the rapid readjustments now under way. What can the summit-goers do to make the promise of prosperity less fragile?

Two things to begin with. The most urgent need is for a reaffirmation and an extension of the Western doctrine that trade protection is an abomination. The West free-traders echo closely St Augustine's plea to be made chaste, but "not just yet". The US government is rightly complaining about European agricultural protection through the Common Agricultural Policy of the European Communities, while protecting both its own farmers and a variety of other declining industries by quota arrangements that conflict with the spirit if not the letter of the General Agreement on Tariffs and Trade. In Europe as well as the United States, advanced industries and their sponsoring governments complain that Japan is selling unfairly in their mutual competition, whence another battery of quotas and "voluntary agreements" to limit imports. Japan is probably right to protest its injured innocence of these charges, but is equally guilty of what may be the most pernicious form of protection in modern circumstances, restraint on the free flow of capital. To ask that these and other restraints on trade should be promptly thrown out of the window would be to ask too much, but to ask that governments should plan eventually to practise what they preach is mere prudence.

The second urgent need is to face up to the reality of international debt. The lenders and the borrowers are at present playing chicken with each other, but everybody knows that the West will not, for example, let Mexico or the Philippines go bankrupt. Politically, these countries are too important, while the collapse of the commercial banks that would be carried with them would cause untold damage in the industrialized world. So



why wait until disaster stares these governments directly in the face? Why not find a way out of the mess before it actually exists? The US administration has shown signs of willing in this direction, although the Secretary of the US Treasury's plan for dealing with international debts (which would have the commercial banks lend more, but with an implicit government guarantee) will not be generally acceptable. A better and more durable plan is needed. That is what Tokyo should be for. □

What's in a name?

The latest name for the AIDS virus is in trouble before the christening is over.

EVERYTHING about Acquired Immune Deficiency Syndrome (AIDS) seems to be controversial. That will be many people's reaction to the report (see p.3) of the arguments that attend the launching this week (see p.10) of the proposed name for the causative virus. The Varmus committee which has laboured for more than a year on the design of a single name to replace the two now in common use will get none of the thanks it deserves for the trouble it has taken. The most serious danger now is that an important field of investigation already too much soured by contending passions will be further embittered by personal considerations. Everybody's interest is that a corrosive rivalry between research groups should be mollified.

The identification of the virus responsible for AIDS is in every sense a remarkable piece of modern history, the significance of which for the treatment of disease is not yet fully appreciated. AIDS as a recognizable disease is barely five years old. Its recognition as an infection, first on epidemiological grounds, is naturally more recent. The first published evidence that a virus is responsible is due to Dr Luc Montagnier, at the Institut Pasteur in Paris, and consists essentially of a number of clinical cases and an electron micrograph including virus-like particles. Soon afterwards, Dr Robert Gallo and his associates at the National Cancer Institute (NCI) of the US National Institutes of Health in Bethesda, Maryland, were able to publish evidence that virus-like particles isolated from patients with AIDS are indeed able to infect other lymphocytes in culture; with his then recent interest in human retroviruses infecting lymphocytes fresh in his mind, Gallo was naturally tempted to think of his virus as one of a series (of which there were then already two). This is how the Pasteur virus and the NCI virus are differently named; Montagnier's is called LAV (for lymphadenopathy-associated virus), Gallo's HTLV-III (for human T-cell lymphotropic virus). Further investigation, which has been breathtaking in its speed, shows the two viruses to be essentially identical, give or take the replacement of a sizeable percentage of the bases in the nucleotide sequence of the virus.

This would not be the first time the same organism has been discovered independently by two different routes. In the past, there has even been confusion of this kind about the naming of chemical elements. Ultimately, of course, there is usually a felt need to make sense of such confusion. What the Varmus committee has been doing, on behalf of the International Committee on the Taxonomy of Viruses, is to force the pace a little, to bring order to a field in which disorder appears rife. The snag is that this laudable attempt should have coincided with a serious dispute between the French and US groups. The two groups have been edgy about each other's proprietary interests in the discovery for some two years. Last year, after techniques for the diagnosis of AIDS antibodies were developed separately in Paris and Bethesda, the Pasteur Institute filed a claim against its US colleagues (or, strictly, against the US Department of Health and Human Services) on the grounds that US reagents were developed with the use of French material. The matter has been made to seem worse (but it may be only an appearance) by the discovery that one of the electron micrographs in Gallo's first paper about the virus was mistakenly an image of the

French virus.

Trying to win agreement on a common name at a time like this is therefore a little like trying to win agreement on a national anthem in the middle of a civil war. It was courageous of the Varmus committee to attempt the task. (A letter from Gallo and his associates explaining how the mix-up happened, and remarking on some of the attendant issues, will appear in *Nature* next week.) Meanwhile, the committee's chances of winning acceptance for its proposals have been diminished by Montagnier's premature announcement of what the Varmus committee had decided. The issue has nothing to do with the recommended name but with the circumstances: bearers of important news are supposed to run on an inside track. It is understandable that Gallo should now be unwilling to use the recommended name for any but generic purposes; he was probably a reluctant member of the naming committee from an earlier stage.

There is nothing much that can be done about all this. It will not make sense to hope that there can be agreement on a common name for the virus that causes AIDS if one or other of the principal investigators in the field should decline to use it. Even if one should be compliant and the other should demur, enforcing commonality would be a disservice. That is why *Nature* will not attempt to lay down the law, but instead seek only to ensure that its readers are not confused. We shall (like our readers) wait and see. Meanwhile, the biomedical research community should do more than it has done so far to resolve the issue that has arisen. The Montagnier and Gallo groups stand out among several working in basic research laboratories that have thrown light on the way that AIDS infections work. Neither has yet cured a patient, but each has given hope to many. In passing, they have given everybody good cause to believe that the investigation of novel infections is now well within the competence of well-organized laboratories. It would be shabby, even disgraceful, if the contribution of these two men and their associates were not widely recognized, even by the principals. That is what the community owes its heroes, or near-heroes: for in this case, that of an infection for which there is still no therapy, the real heroes are the physicians who must deal with patients in ignorance of what may yet be possible. And, of course, the patients themselves. □

Cloud over Chernobyl

The first Soviet nuclear accident to be made public may bring unlikely comrades together.

LAST weekend's nuclear accident north of Kiev is a sobering reminder of the problems of running high-technology enterprises safely. The fact that the first sign of trouble was reported not from the Soviet Union, but more than 1,000 miles away, in Sweden, suggests that the Soviet managers of the plant that released the cloud of activity should have had a better telephone system at their disposal or that there should have been somebody able by other means to tell Sweden about the cloud of radioactivity on the way. Neighbourliness requires no less.

The important question, when the dust settles, is not so much how accidents like these can be prevented, but how we can learn to live with them safely. For, from time to time, even with nuclear plants (which go wrong less often than other kinds of machinery), accidents will happen. There is nothing particularly secret about the Chernobyl plant, whose general characteristics have been declared by the Soviet Union to the International Atomic Energy Agency at Vienna; it consists of six reactors, two still under construction, in which the neutron moderator is graphite and the coolant consists of steam (also a moderator). On paper, such a plant should have a degree of built-in safety. (Let coolant escape, and the chain reaction should be less efficient.) But something went wrong. The Soviet Union's neighbours have a right to know what it was, and also a right to some assurance that the same will not happen a second time. □

Virus nomenclature

Controversy over AIDS virus extends to name

Washington

MORE than a year of patient negotiations to choose a new name for the virus associated with acquired immune deficiency syndrome (AIDS) ended last week in a hectic torrent of letters, leaks and rearranged release dates. Two key members of the naming committee have refused to go along with human immunodeficiency virus (HIV) as a replacement for HTLV-III/LAV, but support for a new name is strong; HIV could well be the eventual choice.

The name change is proposed in a letter published elsewhere in this issue (see p. 10) and simultaneously in *Science*. The letter is the result of a year's deliberations by a specially empanelled subcommittee of the International Committee on the Taxonomy of Viruses (ICTV). Chaired by Harold Varmus of the University of California at San Francisco, the subcommittee consisted of the six members of the ICTV retrovirus study group (John Coffin, Natalie Teich, Kumao Toyoshima, Varmus, Robert Gallo and Max Essex), as well as seven members directly involved in AIDS research or familiar with the field (Ashley Haase, Jay Levy, Luc Montagnier, Steven Oroszlan, Howard Temin, Peter Vogt and Robin Weiss).

By February this year, the committee had tentatively agreed on HIV, and the name was proposed to ICTV's executive committee. David W. Kingsbury, who heads ICTV's human virus subcommittee, says that all 12 of the responses received so far from the 17 executive committee members are favourable to the new name.

But HIV never had unanimous support from Varmus's subcommittee. Nearly half of the members preferred the current compound name, HTLV-III/LAV. Others, including Gallo of the National Institutes of Health (NIH) in Bethesda, Essex of the Harvard School of Public Health and Temin of the University of Wisconsin, preferred human retrovirus (HRV), with numbers assigned sequentially for every new virus. But committee members were anxious to find a name that would eliminate disputes over priority of discovery; for that purpose, HRV was too nonspecific for most committee members.

Armed with a positive response from the ICTV executive committee, Varmus then sent to *Nature* and *Science* a letter proposing the new name, saying a full list of signatories would follow. But the announcement of new human retroviruses late in March dashed hopes for unanimity

for HIV.

The new viruses, called LAV-2 by Montagnier of the Institut Pasteur and HTLV-IV by Essex, may turn out to be variations of the same virus. Montagnier announced his discovery at a meeting in Lisbon before it had been accepted for publication in any journal (see *Nature* 320, 3; 1986). Montagnier's announcement



came slightly more than a week before publication (in *Science*) of Essex's discovery. Essex was incensed by Montagnier's action, and he informed Varmus that he would have nothing to do with the new name, already supported by Montagnier.

Essex says he had never been satisfied with HIV as a name for the AIDS virus. He feels that the name HTLV-III/LAV is so ingrained in the literature that it will not be possible to change it, and calls the committee's efforts "shortsighted". He is also concerned that the new nomenclature would not be appropriate for HTLV-IV, as it does not cause immune deficiency.

Shortly after Essex's decision, Gallo indicated that he would not go along with HIV either. Gallo's attitude hardened further when, two weeks ago, Montagnier delivered a lecture at NIH where he mentioned HIV. A reporter, alerted by Institut Pasteur's New York public relations company, attended the meeting, and a report that HIV would replace HTLV-III/LAV appeared in the *Washington Post* on 19 April.

Gallo was furious. Varmus had decreed that no subcommittee member should discuss the new name with the press before it appeared in a scientific journal. Outraged that Montagnier should have made his announcement "right under my nose", Gallo wrote to Varmus exhorting him to withdraw his letter to *Science* and *Nature*, and saying that he should reconsider the new name.

Varmus was also furious with Montagnier for leaking the new name prematurely, but felt that Essex and Gallo were

being churlish not to go along. After consultation with editors John Maddox of *Nature* and Dan Koshland of *Science*, it was agreed that the letter proposing the new name would appear as soon as possible. A contentious sentence asking editors printing the letter to "urge" authors to use the new name was dropped in the final version.

As for the name itself, even Varmus concedes that HIV has problems, but he believes it represents a reasonable compromise. Several committee members conceded that HIV was not their first choice, but were willing to compromise and to accept it.

David Baltimore, who is heading a study of AIDS for the National Academy of Sciences Institute of Medicine, says the AIDS virus was "crying out" for a name change. Baltimore shares the opinion voiced by many that, because HTLV-III is too dissimilar from its numerical ancestors HTLV-I and II to share the name, a change is needed to eliminate confusion on that point. But supporters of the name HTLV, chosen originally by Gallo in the mistaken belief that it would be closely related to HTLV-I and II, argue that hepatitis A and B are even more dissimilar from each other than are the HTLV viruses, but that those two names coexist without controversy.

Gallo says that he used HTLV-III only after a protocol for naming new human retroviruses had been agreed by an international *ad hoc* committee formed at a Cold Spring Harbor symposium in September 1983. But it is precisely because controversy has arisen over the virus responsible for AIDS, says Kingsbury, that the committee has sought a new name.

Steve Gillis of Immunex Corporation at Seattle, Washington, who is familiar with controversies over new names from his own experience with lymphokines, questions whether a name that is not supported by Gallo can win general support. In addition to Gallo and Essex, a prominent AIDS researcher who asked not to be identified indicated that he will not use the new name.

But support for HIV is already growing among people in the field. Baltimore has already begun to use the name, and Jerome Groopman of Harvard Medical School has included it in a draft of a paper he is preparing.

By the end of the week before publication of the Varmus letter, Gallo was insisting that he would never use HIV when referring to his original isolates, but he might consider using HIV to refer to a "generic" AIDS virus.

Formal endorsement from ICTV for HIV is expected to come in May. But, as ICTV chairman Fred Brown puts it, any name is acceptable "so long as people know what you are talking about".

Joseph Palca

British astronomy

Observatory move stirs passions

THE plan to move the Royal Greenwich Observatory (RGO) from its present location in Sussex to somewhere else in Britain has evoked mixed passions among British astronomers. If possible, feelings run even more strongly than in the high-energy physics community when British membership of the European Organisation for Nuclear Physics (CERN) was being examined more than a year ago. But British astronomers are divided between those who think RGO should stay where it is and those who wish it to be moved to one

quired by SERC, they were told that the requirement of floor space had doubled.

According to a SERC spokesman this week, there is some confusion on this point. He acknowledged that there had been a change of specification, but said that the present requirement for 4,000 m² of space was intended to be a gross figure; usable floor space would be three-quarters of the total. This does not fully meet the Oxford contention that the hospitality the university offers includes a large workshop equipped for work in



or other of the three candidate locations.

The decision that RGO should be moved was taken by the Science and Engineering Research Council (SERC) in March (see *Nature*, 320, 297, 1986). SERC is to decide at its meeting in June whether RGO should, while retaining its identity, go to the Universities of Cambridge or Manchester or, alternatively, unite with the Royal Observatory, Edinburgh (ROE), on its Blackford Hill site.

SERC's announcement of these developments followed an invitation to seven British universities to offer hospitality to a transplanted RGO. Dr E.J.W. Mitchell, SERC's chairman, explained in March that the decision to move RGO had arisen in part from disappointment with the strength of its present relationship with university astronomers. Of the three future locations for RGO, the council said that Edinburgh was the most favoured.

Some of the fiercest reactions to this development have come from universities not now on the short-list of candidate locations. The University of Oxford, one of those invited to provide a site for RGO earlier in the year, is up in arms at what local astronomers consider to have been a change in the rules of SERC's competition when the finishing post was in sight. In a strongly worded protest to SERC, Oxford astronomers have complained that, when they had identified a building capable of providing the 2,000 m² of floor space re-

nuclear physics which will be under-used when SERC ceases to support the tandem electrostatic generator on the site.

The lobbying of the two universities on the short-list is, for Britain, unusually direct. The University of Manchester makes a great deal of its offer of rent-free accommodation, on the site of the Christie Hospital, for RGO. At Cambridge, on the other hand, it seems to be accepted that SERC would have to meet the cost of a building to house RGO, presumably out of the proceeds of the sale of the Herstmonceux buildings: SERC has said that the move will have to be self-financing. In the nature of things, the sale of the castle is at the speculative end of the real-estate spectrum.

Intellectually, Cambridge offers proximity to the Institute of Astronomy and the Cavendish Laboratory, with RGO to be sited in the grounds of the university observatory adjoining the institute. Manchester, by contrast, emphasizes its own close academic links with other universities and polytechnics in the Midlands and northern England. In a letter addressed to all fellows of the Royal Astronomical Society, the Astronomer-Royal, Sir Francis Graham Smith, who is also director of the Jodrell Bank Observatory of the University of Manchester, argues first against the preferred move to Edinburgh on the grounds of geography, the tenuous academic link between the Edin-

burgh observatory and the university and by saying that the continued separate existence of RGO and the Edinburgh observatory offers the best scientific value. He then goes on, after declaring his interest, to put the case for Manchester on the grounds of geography and facilities.

Astronomers elsewhere appear mostly to wish that a decision could be avoided. All speak highly of the work done by the two public observatories in commissioning new telescopes on Hawaii (ROE) and La Palma (RGO); there is particularly high praise for the support offered by RGO to academic astronomers using the 2.5-m Isaac Newton telescope at La Palma.

This argument is lent support by the lack so far of an estimate of the savings that may result from the proposed move, which SERC says is being made on scientific grounds within the constraint that the cost of the move should be supported by the sale of assets. The staff at RGO has already shrunk from its peak of about 240 to a present complement of 190. By the time planned for the move in 1990, it is intended that the observatory's staff should consist of about 100 people in Britain with a further 30 at La Palma.

The opinions of the RGO staff are also divided, but favour Cambridge. A survey carried out last month (to which 55 per cent of the staff responded) showed that 80 per cent would be prepared to move somewhere, of whom 90 per cent gave Cambridge as their preference. Asked to identify the sites to which they would not move, only two per cent of the staff came out against Cambridge. In response to the same question, 60 per cent said they would not move to Edinburgh and 50 per cent dug in their heels against Manchester.

Meanwhile, a regional political lobby is forming to preserve RGO at its present site. The East Sussex County Council and the North Weald District Council, both of which have an interest in the tourists attracted by Herstmonceux, are seeking to influence official opinion, as are local members of parliament.

There is no precedent for political forces such as these influencing the autonomous decisions of research councils, yet SERC will need strong nerves if it is to stick to its timetable of a decision next month. Matters have not been much simplified by reports that Dr Harry Atkinson, the senior official of SERC in charge of astronomy spending (and also chairman of the European Space Research Organisation) had given the members of one committee the impression that there were political considerations in the decision not to vacate the Edinburgh site. Dr Atkinson explained on the telephone this week that he meant only that, in his opinion, SERC should pay attention to the regional distribution of its establishments, but even that may be enough for the Sussex lobby to keep the door ajar. **John Maddox**

SDI in Japan

Growing interest in joint venture

Tokyo

AT the eleventh hour, two Japanese scientific groups have joined forces to organize a signature-gathering campaign across the country in protest against the US Strategic Defense Initiative (SDI). "Physicists and Mathematicians against SDI" claim to have gathered nearly 3,000 signatures from their colleagues at 145 universities and research institutes across the nation. But it looks as though their protest may be too late.

Last week, a report was submitted to the government from a fact-finding mission of government and business groups (see *Nature* 320, 294; 1986). The 55-member mission had spent ten days in the United States visiting SDI-related research facilities and being briefed on the projects' technical aspects. The mission was the third to visit the United States but the first to include representatives from private companies, itself a pointer to the growing likelihood of participation.

Even before the mission left, business leaders had privately admitted that they had been approached by government and that a generally favourable attitude towards participation had developed. Their main worry, one said, is whether the United States would guarantee that SDI would be a long-term project and not susceptible to sudden cancellation. Without that guarantee, some of the smaller companies would find a commitment to SDI research too risky. Given these behind-the-scenes developments, it is no surprise that the fact-finding mission's eight-page report recommends that Japan take part in the "research phase of SDI".

The report describes SDI as being "non-nuclear" and aimed at abolishing nuclear arms, thus making it compatible with Japan's commitment to peace, enshrined in its constitution and in a Diet resolution never to permit nuclear weapons research. It also stresses that SDI is only a research programme and not a plan for deployment of a defence system. But the chief advantages it sees in participation are more practical.

The report says that progress has already been made in the SDI programme and raises the spectre of a "widening technology gap between the United States and Japan". It also asserts that there will be significant spin-offs for the civilian sector including new air-control radars and remote sensors from the surveillance, acquisition, track, kill and assessment (SATKA) section of SDI. The sensors used to detect heat from rocket exhausts will also find applications in medical sensors. Kinetic-energy weapons research will aid aircraft technology with development of new lightweight materials. And

high-power lasers and particle beams can be used in the control of industrial chemical reactions. Altogether, participation will "greatly improve the level of related technologies in this country".

Further consultations will now take place and the final decision made by a group of five cabinet ministers — Defense, International Trade and Industry, Science and Technology, Foreign Affairs and the Cabinet Secretary — along with the Prime Minister. If the answer is positive, as is now expected, they will deter-

mine the form of Japanese participation. The decision is not expected to be in time for the Tokyo summit of the leaders of industrial nations next week.

With the SDI bandwagon already rolling, it is hard to see that a protest from scientists will have much impact. But many of those who have signed the petition opposing SDI are the very people — physicists, mathematicians and computer scientists — who might be expected to be involved in SDI research. They are concerned not only that science is being "increasingly militarized" but also that their research will be diverted into SDI and their publications become restricted "military secrets".

Alun Anderson

Animal protection

Critics assail West German law

Hamburg

THE Bundestag (the West German parliament) has now passed the amended form of the 1972 Animal Protection Act, introduced by the minister of agriculture, nutrition and forestry in the federal government, Ignaz Kiechle. But it seems unlikely that this development will silence the arguments about the use of animals in experiments that have become common in the past few years.

The new law is the first in a series of measures intended to cover several issues in the field of biological ethics. The dispute about animal experiments, involving industrial companies and scientists on one side and anti-vivisectionists on the other, has been especially fierce in recent months. Individual protesters continue to release experimental animals from institute laboratories, but the bombing of a Cologne laboratory last year has not yet been repeated.

On the face of things, the requirements of the new law are tough, and cover all experiments leading to "pain or injury of the animal", a definition that leaves the use of the organs of dead animals exempt from regulation. In future, investigators will have to justify the need for their use of experimental animals by reference to the investigation of the environmental impact of materials on human beings or the risk of new chemical substances used in medicine. The study of human and vertebrate physiology, like basic research in general, will be a valid justification of experiments.

The new law will require that laboratories where experiments with vertebrates are carried out must appoint animal protection officers (*Tierschutzbeauftragten*) who must be qualified physicians, veterinarians or zoologists, and who will be required to supervise the application of the new regulations, such as an interdiction of needless duplication of experiments.

The new law forbids the use of animals in the development and testing of

weapons; animal tests are also forbidden in the testing of tobacco products, washing powders and cosmetics. While the immediate effect of the new law is to restrict animal experiments, there are provisions to encourage the use of alternative methods to the use of animals in research.

Authorization procedures are to be tightened, chiefly by the appointment of a network of commissions to supervise the authorization procedure. The animal protection organizations will have a powerful influence on the proceedings of these advisory commissions, having established the right to nominate up to a third of the members of the commissions.

The argument is certain to continue, if only because its philosophical and legal implications have not yet been talked out. Some jurists argue that the West German constitution (*grundgesetz*) is relevant in that a guarantee of human dignity on the principles laid down by Kant and Schopenhauer cannot also be used to justify the killing of countless animals so as to protect people from over-indulgence.

It is significant that there have been several recent decisions of the constitutional court to the effect that the "absolute" right of freedom of research can be in conflict with novel ethical principles, which may imply that guarantees such as this in the 1949 constitution may be mutable in changed circumstances.

Of the political implications of the new developments, there seems no question. The critics of the new law complain that "those to be controlled are the controllers", and their argument is consistent enough to continue at least until the next federal election in January 1987. The Social Democrats have said that they will amend the law if they win power; both they and the Greens also want the law to be extended to cover the extensive keeping of animals as pets. But scientists consistently oppose the new law as an impediment to research.

Jürgen Neffe

Japanese education

Further iteration towards reform

Tokyo

ATTEMPTS to reform Japan's educational system have taken another halting step forward with the publication last Thursday of Prime Minister Yasuhiro Nakasone's Special Educational Advisory Council's second full-length report.

The report met the barrage of criticism that has greeted every pronouncement of the special council (see *Nature* 319, 347; 1986) and was preceded by the publication of counter-reports from women's groups, teachers groups and unions, proving, as other nations have found, that although it may be hard to effect educational reform, it is certainly not hard to whip up opposition to it. In the new report, for the first time, details begin to emerge of how universities might be reformed.

The central criticisms of the universities are that they are too inflexible, too homogeneous and do not make a contribution to the international research community commensurate with Japan's status as a world power. Furthermore, the universities are resistant to change and do not sufficiently influence educational policy.

The solution to this last problem is seen in the establishment of a "university council" which would have the right of direct access to the Minister of Education, Science and Culture. Its brief would be wide: as well as conducting surveys and collecting information, it would expect to proffer advice on the structure of the universities, their planning and construction, on what areas needed developing to provide trained manpower, the content of university education and the methods of teaching. There are at present a number of bodies representing the universities but none that can speak for all universities — including state and private universities and the junior (two-year) colleges.

To the problems of inflexibility and homogeneity come one familiar solution: the use of a credit system that would allow students to transfer between faculties, or even universities. If necessary, a special institution should be set up to grant degrees to those who have gained sufficient credits from several sources. Inputs to the universities could be made more diverse with quotas set for the enrolment of ordinary citizens and junior college graduates into the four-year universities.

Suggestions for the reform of university research begin with the surprising statistic that the ratio of graduate students to undergraduate students in Japan is the lowest of all the developed countries. In part, of course, this reflects the high number of engineering students who quickly enter industry. Less surprising is the recognition of the lack of technical back-up in the universities: the general rule is that

if you want a piece of equipment you must build it yourself. To bring about reform, young people need better treatment.

The report gives the average age for completion of a PhD course as 27. To make things move faster, it is suggested that it should be made possible to complete a master's course in one year and a doctoral course in three years, with direct entry to a doctoral course after graduation possible for especially talented students.

Postdoctoral fellowships (of which there are still only a few hundred in Japan) should be made much more easily available. And thought should go to the establishment of an independent graduate university for research students. This is something those in the new research institutes for joint use by universities (like those at Okazaki) would particularly welcome, for without attachment to any particular university they have no ready source of graduate students. A graduate university could well find a home at Okazaki.

The role of "joshu", the lowly position of research assistant which is often the first

rung on the ladder to professor but which is sometimes mistaken by professors as a post providing personal attendants, also needs to be looked at afresh, the report says. Low staff turnover is another problem: possible solutions are the usual early retirement, untenured short-term posts and contract research.

Reforms are, of course, expensive, particularly the provision of increased technical assistance, more fellowships, graduate universities and increased foreign exchange. The report points out that 330 thousand million yen will be spent on academic research this year. But estimates of how much reform of postgraduate education would cost suggest that the same amount again would need to be spent over the next ten years. This is unlikely to find favour with the Ministry of Finance, which is already complaining that spending on education is high enough. How high is high enough is hard to measure: the Japan Teachers Union claims that Japan's 1985 education budget accounted for 4.6 per cent of its gross national product compared with 9.1 per cent in Sweden, 8.1 per cent in The Netherlands, 7.2 per cent in the Soviet Union and 6.4 per cent in the United States.

Alun Anderson

Soviet science

Lenin prizes return to basics

THIS year's Lenin Prizes for Science and Technology in the Soviet Union concentrate on fundamental research. Only two of the eight prizes awarded last week have any immediate relevance to the application of research results in production, hailed for the past two decades as the principal target for Soviet science.

Pravda's editorial preamble to the review of the prizes by Academician Anatolii Aleksandrov, president of the Soviet Academy of Sciences, still emphasizes the need to "accelerate scientific and technical progress", but the stress is now on the mobilization of "creative forces" in science rather than on the rapid implementation of results.

All but one of the awards are cumulative, for "cycles" of research and publications going back, in some cases, more than 20 years. The exception is the prize for Yuri Aleksandrov and his team from the Institute of Radio and Electronics of the Soviet Academy of Sciences and the Special Design Bureau of Moscow University, who won the prize for developing a special radar system for mapping the surface of Venus from the Venera-15 and Venera-16 probes. (This is the only prize-winning work for which Academician Aleksandrov mentions spin-off; the system, he says, has also been applied in the study of flooding, irrigation, salination, and soil and water studies in many parts of the Soviet Union.)

The longest "cycle" to be honoured is that of Academician Grigori Devatykh, for his development of methods of obtaining high-purity volatile substances, which he began in 1959. Almost as early, and far more significant, is the starting date assigned to Dr Roman Khesin-Lur'e's work on "molecular principles of the functioning of the genome", which he apparently began to publish in 1960. Lysenkoism was then in full swing, and, according to the Soviet biological establishment, the gene was considered an "empty abstraction", accepted only by reactionaries. It would be interesting to know where Khesin-Lur'e published his work at that time, and what part, if any, he played in the overthrow of Lysenkoism.

Another interesting award is that for "inclusive processes in strong interactions of high-energy elementary particles and the discovery of scale invariance in these processes". Aleksandrov goes out of his way to stress that the new directions opened up by the work done at Serpukhov were afterwards confirmed at CERN and the Fermi National Accelerator Laboratory in the United States, and the team includes not only Soviet citizens but also Dr Nguyen Van Hue, president of the Vietnamese National Centre for Scientific Research. Lenin prizes are normally a purely Soviet affair, but this year there seems to have been a greater emphasis on their place in world science.

Vera Rich

Biology at Berkeley

New order prompts fund-raising

Berkeley, California

THE University of California (UC) at Berkeley has started mailing thousands of glossy brochures to alumni and academic philanthropists in order to raise \$50 million towards a planned \$150 million refurbishment of its biology department. Spurred by the perception that Berkeley has lost pre-eminence in molecular biology and genetic engineering to other UC campuses, the scheme has snowballed into what must be one of the most ambitious reorganizations ever of an academic department.

Hundreds of reports and untold numbers of faculty meetings have resulted in a grand plan aimed at making the most of the new biology by merging and reshuffling most of the seventeen existing de-

partments into three new super-departments based on affinities between research interests. The biggest problem at present is the condition of the life sciences building, completed in 1932 and by all accounts a continual frustration for its inhabitants. Researchers speak of overloaded power lines, frequent floods caused by faulty plumbing and insufficient storage space.

The grand plan calls for two new buildings (one for cell biology and molecular genetics, one for plant and molecular biology, microbiology and biochemistry) and the complete reconstruction of the existing life sciences building (for "integrative biology": organismic botany and zoology, ecology and some genetics, more or less). A special chancellor's advisory committee is making recommendations on all appointments.

Berkeley lacks the tradition of alumni support that private universities such as Stanford have relied upon. Undeterred, the Berkeley campus development office has compiled a table indicating how the required \$50 million might be raised: one donation of \$15 million, two of \$5 million, and so on down to 40 gifts of \$25,000. The state of California has agreed to pay two-thirds of the total cost of the project; overtures will also be made to foundations which have been traditional sources of support for private universities but which have been less generous elsewhere.

Not surprisingly, some professional feathers have been ruffled by the proposed reorganization, often of smaller departments fearful that they might be swallowed up, so disappearing without trace. Biophysics has yet to find a satisfactory home, and is pressing for joint departmental status with cell physiology, for example.

Another problem seems to be that biology at Berkeley is found in two largely autonomous units, the College of Natural Resources (which receives most of its funds from the state as an agricultural experiment station) and the College of Letters and Sciences. The College of Natural Resources, with its applied emphasis, has reservations about throwing in its lot with basic scientists, and some departments in the college (such as entomology) have indicated their desire to retain their independent status, although some of their staff might be affiliated with the new groupings.

Vice-chancellor Roderic Park sees the reorganization as an interim step towards a single new College of Biology (although no such decision has yet been approved) that would encompass both the existing colleges. Park believes that the device of the Experiment Station (which dates back

to the last century) is an inefficient way of supporting research. In his view, California could buy better applied agricultural research if it were to put its grant money up for competitive bids from all comers instead of tying it to a permanent institution which is solely an administrative entity. But such a change would be resisted.

Although there is now general support for the planned changes, many details remain to be sorted out. Curriculum content, for example, has hardly been considered yet. Park says that one result of the reorganization that he would like to see is a better unification of theory and applied development. Wood replies that good researchers will continue to do good research, including collaborations, wherever they are. The debates on reorganization are far from over. **Tim Beardsley**

Engineered vaccines

Instant approval

Washington

THE US Department of Agriculture (USDA) has found a genetically engineered live-virus vaccine to be "pure, safe, potent and efficacious", and has enabled Biologics Corporation of Omaha, Nebraska, the manufacturer, to resume sales after a two-week voluntary suspension.

The suspension came about after charges appeared in the press that USDA had failed to follow proper procedures in approving the vaccine licence. USDA had not notified its own Agricultural Recombinant DNA Research Committee about the application (see *Nature*, 320, 473; 1986). The vaccine, called Omnivac, protects swine from pseudorabies, a potentially fatal disease caused by a herpesvirus. Omnivac lacks a normally present gene for thymidine kinase that allows the virus to enter a latent state in an animal's nervous system, leading to future outbreaks of disease.

USDA's Animal and Plant Health Inspection Service conducted an "environmental assessment" of the vaccine based on available data, but critics of recombinant DNA research are not satisfied, calling instead for a full environmental impact statement. The Foundation on Economic Trends, headed by Jeremy Rifkin, has filed suit in federal court to suspend Omnivac's licence, and a congressional committee is also investigating USDA's procedures in this case.

Meanwhile, the Scientific Advisory Panel to the Environmental Protection Agency (EPA) has issued a favourable report on Monsanto's permit application to test a genetically engineered microbial pesticide in the field. Final EPA approval is expected later this month. **Joseph Palca**

Britain squeezed

THE proposed cuts in French research spending may have a knock-on effect elsewhere and, in particular, on British participation in the European Synchrotron Radiation Facility (ESRF) to be built at Grenoble. The British Science and Engineering Research Council (SERC), which has been keen to participate, has insisted that the principal members of the Grenoble consortium, France, Italy and West Germany, should in return join (and help pay for) Britain's Spallation Neutron Source (SNS) at the Rutherford Appleton Laboratory near Oxford. But now France, the least reluctant of the trio, seems less likely than ever to participate.

According to Pierre Papon, director-general of the Centre National de la Recherche Scientifique (CNRS), which with the Commissariat à l'Energie Atomique (CEA) would have been the leading French partner in SNS, the British have asked potential partners that the costs to Britain of participation at Grenoble should be matched, pound for pound, by European contributions to the planned second phase of SNS development.

"What advantage would there be in that for us?", Papon demanded last week. He is already conscious of the cuts that lie ahead, while CNRS and CEA apparently agree that French investigators already have "quite enough neutrons", and buying a share in SNS would entail an obligation for continuing operating costs.

Although the present ESRF partners would still like to see Britain contribute some 20 per cent of the Grenoble facility, Papon said last week that negotiations are also under way with a consortium of smaller European states, including Austria, Switzerland and Spain. If the negotiations succeed, Papon says, the remaining gap of 10 per cent may be closed without British participation.

Robert Walgate

French budget cut

Chirac takes knife to fatted calf

AFTER five years of socialist expansion, French science is about to go into reverse under the new right-wing Chirac government. In the revised budget for 1986 announced last week, roughly half of all the budget cuts affect science and research. Under the new government, spending on civil research and development will fall by 9 per cent. Since the previous government had planned a 4 to 5 per cent increase in real terms, the result will be a real cut in resources for research and development, the first cut in spending since 1981.

In numbers, FF1,974 million (£84 million) is to be taken from research, most of it from the budgets set aside for the programmes in the gift of the old ministry of research and technology. Funds for the ministry's strategic research in support of new technology (*programmes mobilisateurs*), as well as for other research programmes independent of the research councils, are to be slashed by half, saving FF1,000 million.

Applied research will suffer most, although the chief French programmes in space and nuclear technology are to be protected. ANVAR (Agence National pour la Valorisation de la Recherche), intended to assist technology transfer to small businesses, will be cut by 40 per cent, and the French organization for alternative energy and conservation, AFME (Agence Française pour la Maîtrise de l'Energie), by a third.

The research councils are to lose 10 per cent of their operating budgets, but (with the exception of AFME) the large technology agencies such as the Commissariat à l'Energie Atomique (CEA) will be cut very little if at all. This means that the agencies transferred to the ministry of industry will be left intact but that the organizations left with the rump belonging to the ministry of research and higher education will have been cut.

One small mercy is that funds for the direct support of university research through the ministry of education will also be unchanged, although these are small compared with the sums awarded to universities by the research councils. As yet, the councils have not finally decided how to administer the budget cuts, but the medical council (INSERM) is likely to pass on a 10 per cent cut in laboratory operating costs according to a spokesman last week. The principal research council, the Centre National de la Recherche Scientifique (CNRS), which has to save FF230 million (£21 million), has yet to decide between the deletion of a "whole programme" of research and a general cut of 10 per cent in operating budgets.

The Chirac government is defending the cuts by saying that they are no greater

than the amounts removed from the research budget by the Socialists in 1983, when the previous government was converted from Keynesian to conservative economics. But the government's critics point out that the cuts in mid-1983 were reductions of a budget which had been increased by 16 per cent over that for 1982. Then, there was a real increase of 4–5 per cent in science spending, not the reduction by that amount now in prospect.

M. Alain Devaquet, minister of research in the new government, says that he is "unhappy" with the cuts but that he "understands" the reasons for them. Speaking to journalists in Paris on Thursday last week, he insisted that research remains on the government's list of priorities.

At the same time, Devaquet disarmed

fears that the French research councils CNRS and INSERM would be dismembered, as some members of Chirac's right-wing party were demanding during the election campaign. The minister now says that the research councils have some "small problems" which can be solved without radical change.

The new division of responsibility between the ministries of research and of industry has been partly clarified. Devaquet said last week that his ministry would be left with programmes (now halved in value) intended to encourage industrial research, while the industrial ministry would be responsible for the whole field of industrial development.

This seems to be a victory for those in the industry ministry who have always been suspicious of the upstart initiatives of the research ministry. By the same test, the development is a defeat for the influence of science on French industrial policy.

Robert Walgate

Military research

Campus protests at nuclear tests

Berkeley, California

THE University of California is yet again wracked by protests about military research on its campus at Berkeley. The university administration has now replied to campus petitions alleging failure to oversee the nuclear weapons laboratories it manages for the US government, saying it affirms that work at the laboratories must allow for progress towards nuclear disarmaments as well as design of nuclear weapons.

The issue arises in acute form at the University of California, which has had since the end of the Second World War contracts for the managements of the two principal weapons laboratories in the United States, as well as the Lawrence Berkeley physics laboratory.

The university's present position is spelled out in letters of reply sent to several groups of faculty petitioners who had written to university president David Gardner expressing concern over reports suggesting that nuclear bomb designers at the Los Alamos and Lawrence Livermore laboratories were designing unnecessarily sophisticated nuclear bombs that need constant testing. The charge was made most recently by Hugh DeWitt and Gerald Marsh, physicist at Livermore and at the IIT Research Institute, in the November 1985 issue of the journal *Bulletin of the Atomic Scientists*.

De Witt and Marsh argued that weapons laboratories are now operating under the assumption that continued testing of nuclear weapons will be possible, thus in practice making them necessary and making a comprehensive nuclear test ban impossible. This conflicts with

declared US policy, at least until 1971, when weapons were designed on the assumption that there would eventually be a comprehensive test ban.

The current US administration is the first not to support the idea of a test ban in principle, arguing that continued testing is necessary to ensure the effectiveness and reliability of the weapons stockpile. One petition expressing concern was distributed on the Berkeley campus; its organizers say they believe several hundred faculty members wrote to Gardner on the topic.

The petition argued that any unnecessary testing represented a serious failure of oversight by the university, which manages the laboratories under contract. The university's management contract is voted on by the university every five years, and there have been enough close calls for the administration to be sensitive about allegations such as those of De Witt and Marsh.

Although nobody expects there to be any immediate changes at the weapons laboratories as a result of the university statement, faculty members opposed to the university's role regard it as a step in the right direction. William Frazer, the vice-president for academic affairs at the University of California, says the university is confident that weapons are not being designed deliberately to require nuclear testing and says there are no grounds for thinking the conduct of Livermore improper.

Nevertheless, Gardner is to ask a standing academic review committee to "delve deeper" into the matter, according to Frazer.

Tim Beardsley

Lawrence Berkeley laboratory

SDI brings back old problem

Berkeley, California

THE steep decline in federal support for magnetic fusion research and the concurrent growth in the Strategic Defense Initiative (SDI) budget are posing hard questions at Lawrence Berkeley Laboratory (LBL). Administrators and staff are having to confront again a question many thought had been settled in the early 1970s, when the laboratory's remaining classified research projects were either wound down or transferred to Lawrence Livermore Laboratory, 40 miles away.

Should LBL pursue weapons-related research for the Department of Defense that might, if successful, end up being classified? Recent hard-fought negotiations over publication restrictions on strategic defence research at LBL have now thrown the issue into new relief.

The ability of the national laboratories to set their own policies on classified research was challenged last summer, when the Department of Energy proposed (*Federal Register*, 26 August, p.34662) that they should agree to accept defence work, whether classified or not. That proposal was withdrawn after strong opposition from virtually all quarters, and led LBL director David Shirley to issue a clarifying statement last November to the effect that research at LBL could be "freely communicated to the scientific and technical community". Occasional use of laboratory facilities for classified work by investigators from other institutions would need case-by-case approval in advance by the director.

Similar questions have now been raised

by research, supported by the SDI organization, into neutral-particle beam accelerators and high-current beam instability. The "very objectionable" restrictions on publication originally proposed by the SDI organization have been "fought very hard", according to LBL's general counsel Aleetha Owens. But the fighting is apparently not yet over.

The neutral beam research includes theoretical and experimental work on the development of hydrogen-ion accelerators, and was until recently supported by the Department of Energy's magnetic fusion programme; neutral beams formed by neutralizing hydrogen ions might be a means of heating plasmas to achieve conditions required for fusion. But the magnetic fusion budget at LBL has fallen from around \$6 million to \$2 million in the past two years, and substantial layoffs were likely without an alternative source of funds.

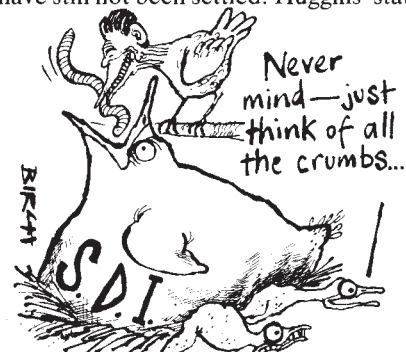
After what sounds like much heart-searching, investigators in LBL's accelerator and fusion research division decided last year to make two applications for SDI funds. A third proposal, for a radio frequency quadrupole accelerator to operate at extremely low temperatures in outer space, was vetoed by the division on the grounds that it had no feasible non-military justification. The result of the applications submitted was a contract for the work on neutral particle beams worth \$2.2 million in the current year, with another \$600,000 for theoretical studies of instabilities in high current beams. SDI also supports work at LBL on infrared sensors.

Laboratory staff say they have been left in no doubt that SDI research results must now be submitted for pre-publication review to ensure that classified information is not disclosed, although the research itself will not be secret. This apparently represents a change of policy by the SDI organization, which insisted last year that there should be no such restrictions on earlier particle beam work. A statement dated 14 February 1986 and signed by James Huggins of the US Army's Strategic Defense Command now circulating among LBL staff says that, following pre-publication review, "any information deemed classified shall not be published".

Staff see Huggins' declaration as drawing a "mighty fine line" between classified and unclassified research. Huggins also states that new interviews and press releases on neutral particle beam work will require advance review for classification and that information that is "non-releasable based on the Militarily Critical Technologies List" shall be revised in consultations with the author.

Owens says she doubts that pre-publica-

tion review would be used to classify LBL research retroactively, but sceptics point to the free electron laser as an example of LBL research that was classified and moved to Livermore when it became militarily important. Although the SDI money at LBL is already flowing for the neutral beam and beam instability research, the official LBL line is that the publication conditions of the contract have still not been settled. Huggins' state-



ment is described as an interdepartmental agreement between the Departments of Defense and Energy which has still not been agreed by the director of LBL.

A further complication is that the University of California, which manages LBL, has endorsed the "no classified research" policy and would take a keen interest in any erosion of that position.

Another concern, according to Jay Marx, assistant director of the accelerator and fusion research division, is the feeling in the division that SDI funds should not exceed 10 per cent of the division's budget because "frankly we're worried about the volatility of the funding". Director Shirley adds that if the laboratory is to undertake a multi-million dollar project, it should have confidence that support will continue for a substantial period, say ten years. But SDI is hoping to achieve breakthroughs by the early 1990s and "that is not research".

There is a conspicuous absence of security controls at LBL; all buildings are open, and visitors can enter the site without so much as a name check by hopping on a shuttle bus at the University of California's Berkeley campus, just a mile down Cyclotron Road. Researchers say they value the unrestricted exchange of ideas with university faculty and others, and are wary of the change in atmosphere that security restrictions would bring about.

Thus James Symons, head of LBL's nuclear chemistry division, says "most people much prefer the present system and hope we can keep it that way". Symons is clearly speaking for the majority in saying that he would react "very strongly and very negatively" to the imposition of restrictions on what he could publish. But the alternative, at least for some researchers, seems to be no research at all. Some might say the objectors had better decide whether they want to have their cake or eat it.

Tim Beardsley

UK flood channel



THE UK Science and Engineering Research Council has this week opened a new flood channel facility built at a cost of £0.75 million at Hydraulics Research Limited, Wallingford. The facility is designed to investigate the interaction between flow in a river channel and over a flood plain; it is 56 m long and 10 m wide, with a discharge capacity of up to $1.1 \text{ m}^3 \text{ s}^{-1}$. □

What to call the AIDS virus?

SIR—The undersigned are members of a subcommittee empowered by the International Committee on the Taxonomy of Viruses to propose an appropriate name for the retrovirus isolates recently implicated as the causative agents of the acquired immune deficiency syndrome (AIDS). Adoption of an internationally acceptable name for this group of viruses has become an important issue because of the widespread interest in AIDS and its origins and because of the multiplicity of names currently in use. Thus the several isolates of what are now evidently closely related members of the same virus group have been called lymphadenopathy-associated virus (LAV), human T-lymphotropic lymphotropic virus type III (HTLV-III), immunodeficiency-associated virus (IDAV), and AIDS-associated retrovirus (ARV). At present, two compound names (HTLV-III/LAV and LAV/HTLV-III) are also used in publications, while the colloquial name, the AIDS virus, is often used by the general press.

We propose that the AIDS retroviruses be officially designated as the human immunodeficiency viruses, to be known in abbreviated form as HIV.

We have considered several issues bearing on this proposal.

(1) The name should conform to common nomenclature for retroviruses, beginning with the host species ("human"), ending with "virus", and containing a word that denotes a major (though not the only) pathogenetic property of the prototypic members of the group ("immunodeficiency"). ("Feline leukaemia virus" and "mouse mammary tumour virus" are two well-known examples of such names for retrovirus species.)

(2) Though the name should clearly link the viruses to the disease with which they are associated, it should not incorporate the term "AIDS", which many clinicians urged us to avoid.

(3) The name should be readily distinguished from all existing names for this group of viruses and has been chosen without regard to priority of discovery.

(4) The name should be sufficiently distinct from the names of other retroviruses to imply an independent virus species, a group of isolates that can presumably exchange genetic information readily with each other but not with members of other known retrovirus species. These other species include the human T-cell leukaemia viruses (for example, HTLV-1 and HTLV-2), which will continue to be named according to a convention adopted by several leading investigators in September 1983. (Though roman numerals are often used to indicate the type of HTLV, arabic numbers were originally prescribed in the agreement and

are thus used here.)

(5) Retroviruses isolated from subhuman primates and found to be genetically related and biologically similar to HIVs should be designated as immunodeficiency viruses of the appropriate host species (for example, simian immunodeficiency virus [SIV] or African green monkey immunodeficiency virus [AGMIV]).

(6) Because HIV isolates are numerous and display considerable genetic heterogeneity, particularly in the *env* gene, it will be necessary for each laboratory to assign subspecies designations to their isolates. We recommend that each laboratory adopt a code with geographically informative letters and sequential numbers to identify their isolates (for example, the 42nd isolate at the University of Chicago could be described as HIV [CHI-42]). Initially, the existing, well-characterized isolates, such as LAV-1, HTLV-IIIB or ARV-2, should be identified as such in publications to ease the transition to a unified nomenclature.

(7) Any future isolates of human retroviruses with clear but limited relationship to isolates of HIV (for example, more than 20 per cent but less than 50 per cent nucleic acid sequence identity) should not be called HIV unless there are compelling biological and structural similarities to existing members of the group.

We hope that this proposal will be adopted rapidly by the research community

working with viruses.

JOHN COFFIN (Tufts University, Medford, Massachusetts, USA), ASHLEY HAASE (University of Minnesota, Minneapolis, Minnesota, USA), JAY A. LEVY (University of California, San Francisco, California, USA), LUC MONTAGNIER (Institut Pasteur, Paris, France), STEVEN OROSZ-LAN (Frederick Cancer Research Center, Frederick, Maryland, USA), NATALIE TEICH (Imperial Cancer Research Fund, London WC2, UK), HOWARD TEMIN (University of Wisconsin, Madison, Wisconsin, USA), KUMAO TOYOSHIMA (University of Tokyo, Tokyo, Japan), HAROLD VARMUS, chairman (University of California, San Francisco, California, USA), PETER VOGT (University of Southern California, Los Angeles, California, USA), ROBIN WEISS (Institute of Cancer Research, Chester Beatty Laboratories, London SW3, UK)

• An earlier version of this letter asked that journals publishing it should make use of the name HIV a condition for the publication of research articles. There is much to be said for the general use of a common name, while the present proposal seems to have much to commend it (except for the possibility of misreading the last two initials as a roman "four"). Nevertheless, *Nature* will continue its present practice of allowing its contributors to use whatever nomenclature seems to them appropriate, while reserving the right to add a clarifying note if this should seem necessary to avoid confusion. Editor, *Nature*.

Disposal of British civil plutonium

SIR—The stated position of the Central Electricity Generating Board (CEGB) on British civil plutonium (*Nature* 318, 406; 1985) has for its keystone an apparent material inaccuracy of very considerable magnitude, namely, the claim that "no plutonium produced for use in weapons has been... exported for use in weapons". Three apparent facts indicate that the claim is false.

First, CEGB is party to a statement that "... approximately four tonnes..." of such plutonium has been exported to the United States under the provisions of the 1958 Mutual Defence Agreement and its 1959 Amendment, Cmd 859 (*The Electrical Power Engineer*, p.163, August/September 1984).

Second, Paragraph C of Article V of Cmd 859 specifies an *exclusive* use of this plutonium: "the preparation or implementation of defence plans". (Colloquially, this appears to mean bomb tests and bombs.)

Third, and setting the issue beyond reasonable doubt, a spokesman for the US government has recently stated that "...the Mutual Defence Agreement actu-

ally requires the United States to use all plutonium obtained from the United Kingdom under the Agreement for military purposes..." (*The Electrical Power Engineer*, February 1985, p38; my italics).

The apparent material inaccuracy is very considerable. Approximately 4,000 kg of plutonium is enough for a thousand bombs, perhaps two thousand if the design is sophisticated, more than enough to devastate the whole of Europe, or the whole of the Soviet Union, or both. Thus, the apparent involvement of the UK "civil" nuclear power programme in weapons production is on a very large scale indeed.

A similar, apparently false statement occurs in a written Parliamentary Answer published in the first weeks of the Sizewell Inquiry: "...no plutonium from the CEGB nuclear programme has ever been exported for use in weapons..." (*Hansard*, Commons, 4 February 1983, col. 206).

R. V. HESKETH

Faculty of Science,
Bayero University,
Kano,
Nigeria

Is there inanimate memory?

It is too easy to forget that pre-Victorian physics ascribed memory to lumps of iron. Now this still surprising phenomenon has a mathematical basis to call its own.

EVERYBODY knows what is meant by hysteresis, chiefly from what school physics teachers have had to say. Take a piece of magnetic material such as iron, put it in a magnetic field whose strength can be varied and, starting with unmagnetized iron, let the strength of the magnetic field increase steadily. To begin with, the magnetization of the iron increases less quickly than the field; anthropomorphically, the iron hankers after its unmagnetized condition. But then the rate of increase becomes greater; magnetization grows on what the economists call an S-shaped curve, the kind of thing that might be represented by a factor such as $[1 - \exp(-x)]$.

The surprise is that the same happens on the way down. Decreasing the external magnetic field gives the impression that the lump of iron hankers after its magnetized condition. Plotting the values for magnetization and external magnetic field on a piece of graph paper, there is a gap between the two S-shaped curves. Many young people are asked to record the difference, but are rarely asked to think what it means. But these days, when school administrators are so concerned with technology, young people may be expected to know that the area of the hysteresis curve is a measure of the energy loss of a transformer using iron as a magnetic inductor.

The more interesting question is that of the physics of the process. How is it that a lump of iron can remember its past history in such a reproducible fashion that its hysteresis curve can be counted as a function of its constitution? To be sure, the area within the hysteresis curve is a function of the frequency with which the cycle is traversed (the faster, the greater the area), which implies that lumps of iron have not merely a memory, but a time-dependent capacity to forget. I.D. Mayergoyz from the University of Maryland has now provided a framework in which to ask how such odd behaviour may arise (*Phys. Rev. Lett.* **56**, 1518; 1986).

Mayergoyz's starting point is in 1935, and the work of the German physicist Preisach who is said (a little disdainfully) to have constructed a model of hysteresis on "some hypotheses concerning the physical mechanisms of magnetization. Abstractly, of course, magnetization followed by demagnetization must entail some kind of energy expenditure. A magnetized state is ordered, and so has

lesser entropy than that from which it was formed. Simple thermodynamics proclaims that achieving such a state in a finite time requires that some work should be done (which will eventually appear as heat). But that is no more illuminating about the underlying mechanism than the calculations economists make to show that a trade surplus (deficit) must always be balanced by an export (import) of capital, whence (in physics) concepts such as magnetostrictive friction, fashionable enough in the 1930s.

That is why the mathematical framework of the discussion may be helpful in its own right. Mayergoyz borrows heavily from the Soviet mathematician Kasnoselski, whom he describes as having distilled from earlier work a way of describing hysteresis. The problem is this. A hysteresis curve is a function in some appropriate space. An input, in the simplest case the changing strength of the external field as a function of time, yields an output characteristic of itself. If the speed of the transition from the unmagnetized to the magnetized condition is immaterial, or is ignored, there remains a problem in that the shape is a function of the initial and final values of the external field. More generally, at any point of the planar graph in which magnetization is represented as a function of external field, there is an infinity of curves that may represent the future behaviour of a lump of iron, each one dependent on a particular past history.

What emerges from this argument is a description very similar to what the statisticians would call a Markov system. The future behaviour (magnetization) of a lump of iron may be determined from a knowledge of its state at some instant and from a knowledge of the magnetic field then and at subsequent times. How could it be otherwise, when causality applies? So there must be something about the initial state that embodies the memory of recent past history. In dynamical systems, it may be the rate at which the system is changing on a microscopic scale (the velocity field); in the relatively static context of magnetization, the pattern of magnetization is the clue to the past.

The important part of Mayergoyz's result is his demonstration that it should be possible to calculate the properties of the operator that transforms the input function of the system (the external magnetic field) by means of a simple series of experiments in which the output

(magnetization) is recorded. The demonstration is mathematical, and in a novel way, but the result is to suggest what the history of mathematical physics has consistently shown to be significant, the forging of relationships between functions of one kind and functions of another. The fact that, in the simplest case of hysteresis, that of magnetism, the results must accord with the thermodynamics of entropy and the explanation may lie in microscopic physics is, for the time being, irrelevant.

One practical implication of the argument stands out. Although a system showing hysteresis will behave in a particular way determined by its starting condition and the external influences upon it, there are general rules that determine the behaviour of systems exposed to extreme conditions that appear to be unaffected by the details of the case. This is mirrored, in Mayergoyz's treatment, by the way in which the local microscopic conditions (say, local magnetic domains), which are the local repositories of memories of the past, will be smeared out or extinguished by extreme conditions.

What may this mean for the concept of memory? The simple and oversimple answer is that there can be no connection between inanimate physical systems such as lumps of iron and biological systems, some even blessed with what is called consciousness, in which memory must be an infinitely more subtle business. But can that in reality be the case? Nobody seriously suggests that human beings, who can remember the past, are on that account exceptions to causality.

On the contrary, it is simplest to suppose that the state of the nervous system at any instant somehow mirrors past experience in such a way as to be able to regenerate images of the past that people call memories. What Mayergoyz has done is merely to show that such a model can represent what happens with inorganic materials that appear to have a memory of the past. That is not surprising in itself. The fun will come if it turns out that the equations explain some of the properties of memory as people know it. And if the physical analogy of magnetization shows that exceptional local patterns tend to be smeared out, what can explain in the human case the enhancement of especially vivid memories? It will be interesting if the tricks that people play on themselves can be derived from theorems of some kind. **John Maddox**

Molecular genetics

Understanding colour vision

from J. D. Mollon

THOSE with deficient or anomalous colour vision live in different subjective worlds from the rest of us. Psychologists have been slow to take an interest in molecular genetics, but two recent papers by Jeremy Nathans and collaborators^{1,2} allow us to trace one detailed mechanism by which specific variations in DNA lead to individual differences in perceptual ability and in subjective experience.

We already know that normal colour vision depends on the presence in the retina of three classes of cone cell, each containing a different photopigment³ (see figure). A fourth pigment, contained in the rod cells, underlies our colourless vision at twilight. Each of these pigments consists of a protein molecule to which is bound a derivative of vitamin A₁, 11-*cis*-retinal. The 11-*cis*-retinal is common to all the pigments, but the pigment absorbs maximally at different wavelengths according to the protein component. We know in some detail how the photopigment, embedded in the infolded membranes of receptor cells, initiates electrical signals when light is absorbed. To work out the wavelengths falling on a local region of the retina, the nervous system must then find the ratio of the quantum catches in different classes of cone. We have a good idea of how this is done^{4,5}. And we know something of how the local ratio is compared with those measured elsewhere in the visual field, so as to give us accurate colour perception despite changes in the spectral composition of the illumination⁶.

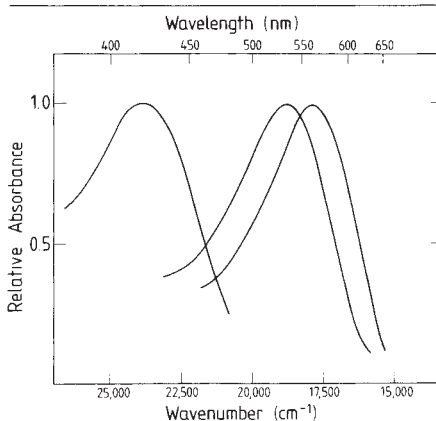
A major addition to our knowledge is provided by the new reports of Nathans *et al.*, who isolated and sequenced the genes that specify the three proteins underlying normal vision¹ and show how the genes are altered in cases of colour deficiency². Nathans *et al.* began with a radioactively labelled fragment of DNA derived from the gene that specifies the rod pigment of cattle. With this probe they were able to identify in human DNA not only the rather similar gene specifying the human rod pigment, but also — by manipulating the experimental temperature and thus the readiness of DNA to 'hybridize' with a partially mismatching probe — they were able to identify other, rather less similar, genes that were likely to be those that specify the cone pigments.

To localize genes on particular chromosomes, probes were prepared from each of the genes and were hybridized with DNA from a panel of mouse-human hybrid cells that retain various subsets of the human chromosomes. From the pattern of re-

sults, the gene for the rod pigment is localized on chromosome 3 and one of the putative cone genes is localized on chromosome 7; the latter gene was taken to be that for the short-wave pigment because disorders of this pigment are known not to be sex-linked. The other genes lie on the q-arm of the X-chromosome. To distinguish genes for the middle- and long-wave pigments, Nathans *et al.* examined 25 colour-deficient men, collating subjective measurements of colour vision with the pattern of hybridization to probes derived from the X-chromosome genes. One curious finding was that more than one middle-wave gene might be present on the X-chromosome of a colour-normal man. But there was never evidence for more than one long-wave gene.

Evolution of colour vision

Why are inherited anomalies of the short-wave cones very rare whereas anomalies of the middle- and long-wave receptors are so very common? Why are the absorbance curves of our photopigments (see figure) so unequally spaced in the spectrum? Why is our visual resolution in



Absorbance spectra of the three cone pigments of human vision. These curves are plotted from Table 2 of ref. 3 and are based on microspectrophotometric measurements of cone cells from human retinas. Peak sensitivities lie in the violet (short-wave), green (middle-wave) and yellow-green (long-wave) regions of the spectrum.

space and time so poor when it depends only on signals originating in the short-wave cones of the retina? These and other features of human vision can be explained by an evolutionary scheme that has recently been suggested by various lines of evidence and is now supported by the work of Nathans *et al.*

Widespread among mammals is an ancient, dichromatic form of colour vision that depends on a comparison of the quan-

tum catches in a short-wave class of cones with peak sensitivity near 430 nm; and in a second class of cones with peak sensitivity varying between species but usually in the range 510 to 570 nm (ref. 7). The short-wave cones are sparse^{3,8,9} and are probably used only for colour vision, whereas all the other business of daytime vision (such as spatial discrimination and the detection of movement or flicker) depends on the second, more numerous, class of cones. The explicitly chromatic information — provided by the ratio of quantum catches in the two kinds of cone — is carried within the visual system by a sub-channel that seems to be electrophysiologically and morphologically distinct^{10,11}.

No doubt a trichromatic, or three-variable, system of colour vision has more than once appeared in the mammals. But our own form of trichromacy seems to be shared only with the Old World monkeys and the apes¹². We, and these primate relatives, enjoy a second dimension of colour vision, provided by a comparison of the quantum catches of the 530- and 560-nm cones. This comparison allows us to make fine discriminations in the red-green part of the spectrum. A distinct retinal channel may not yet have evolved to handle this second type of chromatic information. It is carried by a class of retinal ganglion cells that also extract spatial contrast⁴.

The very recent differentiation of the middle- and long-wave photopigments has probably been achieved by duplication of a single gene on the X chromosome. Nathans *et al.*^{1,2} confirm that the genes for the middle- and long-wave pigments lie in the same region of the X chromosome and they show that 96 per cent of the inferred amino-acid sequence is identical for the two. By contrast, the short-wave pigment shows only a 43 per cent identity with the middle- and long-wave pigments. This is comparable with the degree of identity between the latter pigments and the human rod pigment.

Colour blindness

About 8 per cent of males in Caucasian populations inherit some form of colour deficiency or anomaly. About 2 per cent are dichromats, requiring only two primary wavelengths to match all colours. From reflection densitometry of the living fovea¹³, and from microspectrophotometry of individual receptors¹⁴, there is direct evidence that dichromats lack one of the three cone pigments. Six per cent of Caucasian males are anomalous trichromats, who, like normal observers, require three variables in a colour-matching experiment but who make matches that are different from the normal¹⁴. The most favoured explanation of anomalous trichromacy is that one of the three photopigments is shifted in its spectral position.

What are the genetic abnormalities that

underline dichromacy and anomalous trichromacy? Visual scientists have probably entertained rather simple notions of these disorders, assuming that local mutations of the gene lead to local substitutions in the sequence of amino acids. Nathans *et al.* show that such point mutations are unlikely to be a common source of colour deficiency. Rather, they suggest, the trouble lies in the juxtaposition and the similarity of the genes for the middle- and long-wave pigments. During the formation of ova and sperm cells, when crossing-over occurs between members of a pair of chromosomes, misalignments of the DNA may occur in the region that encodes the middle- and long-wave pigments. Because the sequences are so similar, the long-wave gene on one chromosome might get itself apposed to a middle-wave gene on the other chromosome. If now crossing-over occurs and if the breakpoint lies in the duplicated spacers between genes, then one chromosome might lose its middle-wave gene and the other chromosome end up with two. If, on the other hand, the breakpoint occurs within a gene, a hybrid gene might be formed that consists of part of the long-wave gene and part of the middle-wave gene.

A man who inherits an X chromosome that lacks the middle-wave gene will be a dichromat, but if the X chromosome carries a hybrid gene, then the outcome will depend on where the breakpoint occurred at the time of crossing-over. There is one sub-region of the gene that seems to

determine the actual spectral sensitivity of the resulting photopigment. If this sub-region is drawn from the long-wave gene, then the hybrid gene may produce a photopigment very similar to the normal long-wave pigment; the owner of this chromosome will again be a dichromat. But there may be a whole range of possible breakpoints and some of these may produce hybrid genes that encode photopigments with spectral sensitivities intermediate between those of the normal long- and middle-wave pigments; a man who inherits such a hybrid gene will be an anomalous trichromat. □

1. Nathans, J., Thomas, D. & Hogness, D.S. *Science* **232**, 193 (1986).
2. Nathans, J., Piantanida, T.P., Eddy, R.L., Shows, T.B. & Hogness, D.S. *Science* **232**, 203 (1986).
3. Dartnall, H.J.A., Bowmaker, J.K. & Mollon, J.D. *Proc. R. Soc. B* **220**, 115 (1983).
4. Gouras, P. *Progr. Retinal Res.* **3**, 227 (1984).
5. Boynton, R.M. *Human Color Vision* (Holt-Rinehart-Winston, New York, 1979).
6. Zeki, S.M. *Nature* **284**, 412 (1980).
7. Jacobs, G.H. *Comparative Color Vision*, (Academic, New York, 1981).
8. Mollon, J.D., Bowmaker, J.K. & Jacobs, G.H. *Proc. R. Soc. B* **222**, 373 (1984).
9. Sperling, H.G. in Cool, S.J. & Smith, E.L. *Frontiers in Visual Science* (eds Cool, S.J. & Smith, E.L.) 138 (Springer, New York, 1978).
10. Gouras, P. & Zrenner, E. in *XXIII Concilium Ophthalmologicum, Kyoto* (ed. Shimizu, K.) 379 (North-Holland, Amsterdam, 1978).
11. Mariani, A.P. *Nature* **308**, 184 (1984).
12. Rushton, W.A.H. *Nature* **206**, 1087 (1965).
13. Mollon, J.D., Bowmaker, J.K., Dartnall, H.J.A. & Bird, A.C. in *Colour Vision Deficiencies VII* (ed. Verriest, G.) 303 (Junk, The Hague, 1984).
14. Rayleigh, Lord *Nature* **25**, 64 (1881).

J.D. Mollon is a Lecturer in the Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK.

Taxonomy

Taxonomic tangles from Australia

from Tony Thulborn

IN THE next few years, many of the world's biologists may find themselves beset by taxonomic problems originating from Australia, where a pair of taxonomists is engaged in upsetting the stability of zoological nomenclature. So far, the pair have confined their attention to the Australian herpetofauna, which has been augmented by no fewer than 577 questionable taxa in the past 18 months. As a result, the comparatively stable nomenclature of Australia's reptiles has threatened to collapse into near chaos. Readers will be alarmed to know that the perpetrators now plan to turn their attention to the rest of the world's reptiles.

The saga began in 1981, when R.W. Wells undertook to edit a new periodical, *Australian Journal of Herpetology*, on behalf of the Australian Herpetologists League. Manuscripts were solicited on the understanding that they would undergo rigorous review by an editorial board comprising three renowned herpetologists. The first two issues of *Aust. J.*

Herp. appeared on schedule, and the journal began to draw subscriptions from individuals and libraries throughout Australia and overseas. Things started to go wrong after Wells left his undergraduate studies at the University of New England (Armidale, New South Wales) and moved to Sydney. *Aust. J. Herp.* failed to appear for the next two years, although Wells maintained an address at New England and continued to receive manuscripts.

In 1984 Wells unveiled the third issue of *Aust. J. Herp.* It contained only one paper — a 56-page 'Synopsis of the Class Reptilia in Australia', by Wells and his co-author C.R. Wellington. Manuscripts that had been refereed and accepted for publication in order of receipt had been pushed aside for the Wells and Wellington paper; in fact none of the manuscripts processed by the editorial board has yet been published, although information from at least one of them was transplanted into the Wells and Wellington 'Synopsis'. Authors requesting the return of their manuscripts

have so far met with no success. The third issue of *Aust. J. Herp.* revealed another surprise: its copyright had been transferred to an organization called 'Australian Biological Services', with Wells as 'Managing Editor' and Wellington as 'Advertising Sales Manager'.

The 'Synopsis' itself can only be described as shoddy work — there are no illustrations, and the diagnoses of taxa are so perfunctory as to be useless. Nevertheless, Wells and Wellington designated 33 new genera and 14 new species and elevated 208 synonyms or sub-specific names to the rank of species. The reactions of professional herpetologists ranged from disbelief to outrage. Most dismayed of all were those who found that their casual remarks about their research findings had been published as definitions for new taxa. At its annual general meeting in 1984 the Australian Society of Herpetologists discussed the possibility of suppressing the Wells and Wellington nomenclature, and several individual workers expressed their opinions in letters to *Herpetological Review* (**16**, 4; 1985).

Wells and Wellington were not deterred. In a 98-page publication dated 1 March 1985, identified as *Aust. J. Herp. Supplementary Series No. 1*, and titled 'A Classification of the Amphibia and Reptilia of Australia' they update their previous synopsis of the reptiles and proceed to treat the frogs in similar fashion. The document enumerates 66 new genera (9 of them resurrected from synonymy) and 256 new species (119 of them as resurrected synonyms). It also claims to designate lectotypes for 108 species.

In the bibliography to this second paper there is a list of more than 500 papers said to have been published by Wells in *Australian Herpetologist* since 1983. The existence of those papers is questionable: *Australian Herpetologist* is a 'phantom' publication that cannot be located by librarians, and Wells has not responded to requests for copies of his papers. Also mentioned in the bibliography is a 10-volume work on 'The Herpetology of Australia' by Wells and M. Peterson; this is stated to be 'in press' under the imprint of Australian Biological Services.

Wells and Wellington have created much confusion. Thanks to the convenience of modern word-processing equipment they have been able to exploit existing databases (for example, *The Zoological Catalogue of Australia* Vol. 1, Reptilia and Amphibia) and to re-shuffle them into chaos. The resulting taxonomic work is hopelessly inadequate. Names have been coined for animals on the grounds that they were mentioned by previous authors as local variants. Other taxa are based on nothing more than a reference to a photograph published elsewhere. Type specimens are not identified (unless their registration numbers appear in existing

publications). In several instances there is an attempt to designate a lectotype from an unseen series of syntypes by the ploy of nominating the 'large specimen'. More than 300 names have been elevated to species rank without any explanation or definition beyond "We herein resurrect from synonymy . . .".

The choice of formal names is odd, to say the least. For example, Wells and Wellington introduce the genus names *Licentia* (meaning 'freedom to do as one pleases'), *Phthanodon* ('to anticipate' or 'to do first'), *Contundo* ('to crush') and *Solvonemesis* ('to set free the goddess of retributive justice'). Then there are names that can only be regarded as frivolous: the bestowal of a patronymic species on the Australian Prime Minister, the naming of a species for Telecom Australia, and *Vaderscincus* ('named for Mr Darth Vader'). The name *Erotoscincus* needs no comment.

While the taxonomic practices of Wells and Wellington would not be tolerated in conventional publications, it is not absolutely clear that they contravene the rules and recommendations of the Inter-

national Code of Zoological Nomenclature. Unfortunately, it seems certain that some of their name-games will amount to technically valid combinations.

What is to be done? First of all, biologists outside Australia should be alerted to the situation. Wells and Wellington have placed themselves beyond the pale of legitimate science, and their publications and taxonomic revisions should be treated with extreme scepticism. The Australian Society of Herpetologists plans to approach the International Commission on Zoological Nomenclature with a view to seeking formal suppression of the Wells and Wellington nomenclature. However, such a procedure could be expensive, protracted and somewhat uncertain in its outcome — especially if the Commission refuses 'blanket' suppression and examines each of the 577 cases on their individual merits. Meanwhile, it may suffice that taxonomists are aware of what has been happening to the nomenclature of Australian herpetology. □

Tony Thulborn is in the Department of Zoology, University of Queensland, St Lucia, Brisbane, Australia 4067.

Ocean Drilling Program

Mid-Atlantic bare-rock drilling and hydrothermal vents

from the Leg 106 shipboard scientific party*

LEG 106 of the Ocean Drilling Program (ODP) has taken two significant steps in the study of geological processes at mid-ocean ridges. The first essentially 'zero-age' crustal hole was drilled on the plateau at the summit of a small axial volcano in the Mid-Atlantic Ridge Rift Valley 2,200 km south-east of Bermuda; and engineers demonstrated the feasibility of 'bare-rock', spud-in drilling and coring. The survey of a hydrothermal area in the rift valley also led to the discovery of an active vent field with large sulphide chimneys, hot 'black smoker' vents and a unique biological community. Several drill holes in this area, the first ever drilled in an active, submarine hydrothermal system, recovered massive sulphides (Fe and Cu) and unconsolidated hydrothermal deposits of Zn, Cu and Fe sulphides. Despite the need for technical developments to improve drilling rates and core recovery in these difficult conditions, Leg 106 opens

up exciting possibilities for crustal drilling at oceanic spreading centres.

In submarine volcanic areas, cold sea-water circulates close to hot, intrusive rocks deep within the crust where it is heated to very high temperatures and upwells as hot springs on the sea floor. Although at least six active hydrothermal vent areas have been identified along medium- and fast-spreading ridges in the eastern Pacific, until recently no direct evidence for such activity was known along the slow-spreading Mid-Atlantic Ridge. During a pre-drilling site survey in May 1985, W.B.F. Ryan (Lamont-Doherty Geological Observatory) and L. Mayer (Dalhousie University) took photographs of pockets of greenish-white, mottled hydrothermal sediment, crabs and small, spindly worms indicative of recent hydrothermal activity in an area about 25 km south of the Kane Fracture Zone. The new hydrothermal vent field was discovered during a 36-hour video survey of this area on ODP Leg 106. The sulphide chimneys range up to several metres in diameter and more than 11 m in height, and display the spectacular dendritic, tubular structures and elaborate ornamentation previously seen on the East Pacific Rise and Juan de Fuca Ridge.

Ten shallow holes were drilled at this

site to sample the hydrothermal deposits and the underlying basement rocks, two of which are located at the foot of a large black smoker. The hydrothermal deposits are surprisingly thick — more than 13 m at the foot of the black smoker and at least 3–6 m more than 17 m away from the vent. Recovery of the sand-sized sulphide deposits was generally poor, but our analyses indicate that they are composed mainly of chalcopyrite (CuFeS_2), sphalerite (ZnS), pyrite (FeS), marcasite (FeS_2) and pyrrhotite (Fe_{1-x}S) in varying proportions. Lenses of massive (Fe + Cu) sulphide were also drilled and recovered from a hole at the base of the black smoker. Where recovered, the underlying basalts were fresh.

The vent field is also associated with a diverse biological community that is surprisingly different from that observed in the Pacific. We did not observe any clams, mussels or large, sessile tube worms near the vents. The most common forms of life are large swimming crustacea, small shrimp-like organisms and long (30–60 cm), flat, snake-like swimmers (possibly eels) that give the area the name of the Snake Pit. The shrimp-like organisms seem to congregate on rock surfaces near the vent orifices while sessile little white balls (possibly some type of anemone) litter the sea floor around the vents. In general, the community appears to consist of smaller, more mobile organisms than those previously found at submarine hydrothermal vents.

Previous attempts to drill to significant depths in young oceanic crust have been unsuccessful because of difficult drilling conditions in the fresh, abrasive and highly fractured basaltic rocks found at

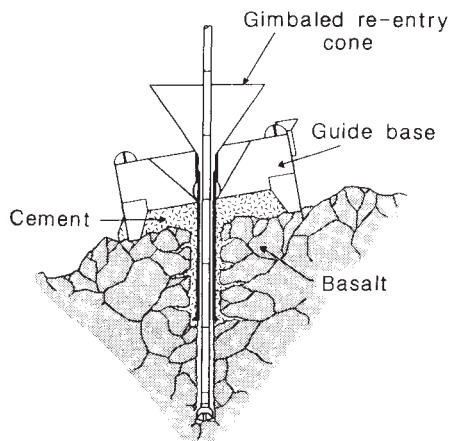
THE three-man deep-sea submersible *Alvin* is going to take a close look at some hydrothermal vents, but this time in the waters of the Atlantic rather than the Pacific Ocean, where all the vents have so far been found. These sites where hot brines emerge from the sea floor, together with their characteristic 'black smokers', sulphide chimneys and peculiar ecology, have now been found on the Mid-Atlantic Ridge by two groups. On page 33 of this issue, Rona *et al.* describe vents found by dredging and video cameras during a cruise in July–August 1985, and on this page, the ODP Leg 106 scientists reveal similar finds made quite close by in December of last year. The level of excitement generated by these discoveries can be judged by the fact that Rona has managed to get time to use *Alvin* in the early summer of 1986. Normally he would have had to wait until at least 1988 to stand a chance of getting a slot on a cruise. But the *Alvin* team just happened to be planning a visit to the mid-Atlantic, and a few extra days have been found to look more closely at the hydrothermal vents.

Peter Gambles

*Co-chief scientists: Robert S. Detrick (Univ. Rhode Island, Kingston) and Jose Honnorez (Univ. Louis Pasteur, Strasbourg, France and Univ. Miami). ODP Staff scientist, Andrew C. Adamson (Texas A & M Univ.). Also Garret W. Brass (NSF), Kathryn M. Gillis (Dalhousie Univ.), Susan E. Humphris (Woods Hole Oceanographic Inst.), Catherine Mevel (Univ. Pierre et Marie Curie), Peter S. Meyer (Woods Hole Oceanographic Inst.), Nikolai Petersen (Univ. Muenchen), Martina Rautenschlein (Max-Planck-Inst. fur Chemie), Tsugio Shibata (Okayama Univ.), Hubert Staudigel (Scripps Inst. Oceanography), Anita Wooldridge (Univ. Miami) and Kiyohiko Yamamoto (Tohoku Univ.).

mid-ocean ridges. The requirement for significant thicknesses of sediment (>100 m) on the basement to spud-in to the crust also precluded any drilling within the narrow, zero-age crust of the accretionary zone itself.

The main objective of ODP Leg 106 was to test and evaluate new bare-rock drilling techniques at the site in the Mid-Atlantic Ridge Rift Valley and to establish a zero-age crustal hole that could be re-occupied and deepened on subsequent legs. Among the new engineering systems tested were a hard-rock guide base to confine the drill bit during initial spud-in on sediment-free rock (see figure); a low-light television camera system for imaging the sea floor and monitoring drilling operations; and



The hard rock guide base used on Leg 106, designed to provide the lateral support necessary for the drill bit to spud-in to hard, volcanic rock on slopes up to 20 degrees and water depths of up to 6,000 m. The 20-ton guide base stands on four legs and is 5.7 m square and 2.3 m high. It is lowered to the sea floor and cemented in place with 56 m³ of cement. A cone 5.3 m in diameter and 2 m deep is built into the base and serves as a temporary re-entry cone. Once the hole is started and cased, a permanent, gimbaled re-entry cone is positioned in the base.

new downhole drilling and coring motors to facilitate bare-rock spud-in.

The site chosen for the first bare-rock drill hole is about 70 km south of the Kane Fracture Zone on the flat summit plateau of a small axial volcano. This site is characterized by bulbous and tubular pillow lavas up to 1–2 m in diameter with a light dusting of pelagic sediment. The hard-rock guide base was successfully lowered 3,344 m to the sea floor, released and cemented without incident. A positive-displacement, downhole drilling motor was used to spud-in to the basaltic crust. The drilling motor and conventional rotary drilling advanced the hole to a total depth of 33.3 m below the sea floor before drilling was terminated after 25 days.

The average recovery for the 26.7 m of cored hole was 23 per cent, comparable with that experienced during previous

drilling in young volcanic rocks. All the recovered rocks are very fresh, plagioclase-olivine, sparsely phryic basalt. The texture of the groundmass ranges from glassy to subvolcanic to intersertal to intergranular, indicating that most samples are probably derived from parts of pillow

lavas. The presence of plagioclase and olivine glomerocrysts and the absence of chromian spinel suggest that the basalts are typical, moderately evolved mid-ocean ridge basalts. The drill hole was left cased and open so that it can be deepened when the ship returns later this year. □

Adrenergic receptors

Some similarities and surprises

from Peter Newmark

Two particularly striking results have emerged from the sequence, reported by Richard Dixon *et al.* on page 75 of this issue, of the gene for one of the most studied of all receptors, the β -adrenergic receptor through which adrenaline and other catecholamines exert their physiological effects. The first is that there are some structural resemblances between the receptor and rhodopsin, the protein at the front end of the machinery in rod cells of the retina that converts light into the visual response. The second surprise is that the protein-coding sequence of the gene that encodes the hamster β -adrenergic receptor is not interrupted by non-coding sequences (introns).

Adrenergic receptors have been subject to the relentless prying of endless ranks of scientists. Physiologists have established the responses to catecholamines mediated by them. Pharmacologists have categorized them into types (α_1 , α_2 , β_1 and β_2) and established their behaviour (for example, desensitization and down-regulation, to use the operational terms beloved of pharmacologists) in response to various types of exposure to catecholamines. And biochemists have elucidated the initial events that follow the binding of catecholamine to the receptor.

What biochemists (biochemical pharmacologists?) have established is that catecholamine binding to β -adrenergic receptors induces their coupling to so-called G (or N) proteins (which come in several types, each of which has three subunits, α , β and γ). The coupling is followed by the binding of GTP to the α -subunit of the G protein, which then dissociates from its partners to activate adenylate cyclase, and thus the synthesis of cyclic AMP. This is a transient event, lasting only until the α -subunit has hydrolysed the GTP that is bound to it. The cyclic AMP produced during that time mediates the eventual physiological responses to the catecholamine.

Dixon *et al.* report two ways in which the structure of the β -adrenergic receptor protein, whose amino-acid sequence they deduce from the DNA sequence, resembles rhodopsin. First, there are three peptides within the receptor sequence that have a considerable number of amino

acids in common with peptides in roughly equivalent positions within rhodopsin; and second, rhodopsin and the receptor have "remarkably similar" hydropathicity profiles (providing some measure of which parts of a protein lie within the cell membrane and thus how the protein is threaded through the membrane).

These structural resemblances make sense in the context of biochemical parallels between the two systems. When rhodopsin is stimulated by light, it couples to transducin, a G protein whose α -subunit binds GTP and, until the bound GTP is hydrolysed, activates a phosphodiesterase that in turn hydrolyses cyclic GMP. The resulting closure of ion channels on the rod cell membrane produces the visual response (see Altman, J. *Nature News and Views* 313, 264; 1985).

It is because of the structural resemblances between the two proteins that it comes as a particular surprise to find that the gene for the β -adrenergic receptor contains no introns. In any case, it joins a very select band: histone and interferon genes are the only other mammalian genes to lack introns; and introns have so far turned up in each of the genes for cell-surface receptors. In the particular case of the rhodopsin gene and its relatives, there are four introns in the mammalian gene, four or five in the genes that encode the three human colour pigments (Nathans, J., Thompson, D. & Hogness, D.S. *Science* 232, 193; 1986) and three or four in two different opsin genes of the fruitfly (see Cowman, A.F., Zucker, C.S. & Rubin, G.M. *Cell* 44, 705; 1986). Dixon *et al.* are confident that they have sequenced the real gene for the receptor rather than a pseudogene, so their discovery will be interesting grist for the mill of those who try and explain the evolutionary goings (and comings?) of introns.

And for those with an interest in β -adrenergic receptors either as a prime example of a receptor or as a target for drugs that modify the physiological responses to catecholamines, the structure of the receptor itself, quite apart from its resemblance to rhodopsin, should prove immensely valuable. □

Peter Newmark is Deputy Editor of Nature.

Astrophysics

The most reliable black hole yet

from M.G. Watson

BLACK holes are frequently invoked by astronomers to explain the vast amount of energy radiated from the nuclei of many galaxies. The energy is produced, in this picture, as material surrounding the black hole falls down its extremely steep gravitational potential well. The same phenomenon is believed to occur on a much smaller scale in a few X-ray-emitting binary stars in or near our Galaxy, providing the best available evidence that these theoretically predicted objects exist at all. In these X-ray binaries the infalling matter eventually accreted by the black hole is supplied by the companion star. A recent investigation of a system of this type — an X-ray transient — has produced results which arguably place the existence of black holes on considerably firmer ground (McClintock, J.E. & Remillard, R.A. *Astrophys. J.* in the press).

X-ray transients are a small but important subclass of X-ray binaries. Most are believed to be systems containing a faint red star and an accreting compact object (a neutron star or black hole). As such they are closely similar to the persistent low-mass X-ray binaries (White, N.E. & Mason, K.O. *Space Sci. Rev.* **40**, 167; 1985), the principal difference being that X-ray transients are only luminous X-ray emitters during their relatively short (of the order of months) outbursts which recur on timescales of years, a pattern of behaviour which presumably reflects a highly variable mass transfer rate onto the compact object. The X-ray transient A0620-00 is perhaps the most famous member of its class as it was, for a few months in 1975, the brightest celestial X-ray source ever recorded (Elvis, M *et al.* *Nature* **257**, 656; 1975). Despite the enormous observational attention that it attracted then, little was learnt about the properties of the underlying binary system. A decade later, the new detailed optical study of the A0620-00 system (now in its quiescent state) by McClintock and Remillard provides compelling evidence that the binary system contains a black hole.

McClintock and Remillard's result is based on several years of photometric observation of A0620-00 in quiescence that provide a unique orbital period of 7.75, together with just one night's spectroscopy which allowed them to determine the radial velocity curve of the faint red companion to the compact object by measuring the periodic Doppler shifts of its absorption lines. The amplitude of the radial velocity curve, $457 \pm 8 \text{ km s}^{-1}$, is very large for a binary with such a short

period, and this measurement alone sets a firm lower limit to the mass of the compact object of 3.2 solar masses. This is significantly above the maximum mass for a neutron star for the stiffest equation of state of neutron matter considered realistic, leading inexorably to the conclusion that the compact object must be a black hole. A more realistic estimate of the compact object mass can be made by assuming a reasonable mass for the red companion star, and using the constraints on the orbital inclination placed by the lack of X-ray eclipses and the modelling of the photometric light curve. On this basis McClintock and Remillard conclude that the mass of the compact object is likely to be greater than 7 solar masses.

The X-ray transient A0620-00 thus joins CygX-1 and two systems in the Large Magellanic Cloud, LMC X-1 and LMC X-3, as the fourth system which is a good candidate for a black-hole binary on the basis of dynamical measurements of the compact object mass. In contrast to the other three black-hole binaries which have hot, luminous, massive companions, the companion in A0620-00 is clearly a faint, low-mass red star. The presence of a massive black hole in this system may pose a fresh set of evolutionary questions. In particular it is difficult to understand how the black hole could have formed in this system without the binary being totally disrupted. The mass of the progenitor star would have had to be at least 20 solar masses (probably more), hence it seems very likely that the formation of the black hole would have involved the loss of more than half the system mass, leading to the binary becoming unbound. The present short binary period is also difficult to reconcile with the size of the original binary system required to accommodate the large, massive progenitor of the present black hole. The difficulties associated with understanding the evolutionary history of A0620-00 suggest that it is a rather rare system. It is too early to say whether this is indeed the case, because the binary system parameters of a large fraction of all X-ray binaries, transient or otherwise, have yet to be measured, but it is interesting to note that the empirical X-ray spectral classification scheme proposed by White and F.E. Marshall (*Astrophys. J.* **281**, 354; 1984) suggested several new black-hole candidates, one of which was A0620-00.

In the case of A0620-00 there can be little doubt about either the correctness of the radial velocity measurements or their interpretation. One of the standard sceptical responses to any unlikely dynamical

measurement has been to suggest that the system is a triple, rather than a binary. This argument cannot be applied to A0620-00, given the small size of the system implied by the 8-hour period. Thus, A0620-00 is arguably the most reliable (or the least unreliable) black-hole candidate yet discovered. Because of this, and because it is a nearby galactic system, it may be that the impact of this result will be mainly psychological in making the possibility, if not the reality, of black holes of stellar mass seem more appealing to many astronomers. □

M.G. Watson is in the X-ray Astronomy Group, Department of Physics, University of Leicester, Leicester LE1 7RH, UK.



100 years ago

New System of Earthquake Observations in Japan

OWING to the invention of new seismographs by the members of the Seismological Society of Japan, there has been of late a complete change in the system of earthquake observations in this country. The Meteorological Bureau now employs the horizontal pendulum and vertical-motion seismographs of Profs. J. Milne and T. Gray, and of Prof. J.A. Ewing for systematic observation, while the Imperial University of Tokio publishes from time to time detailed accounts of particular and more interesting shocks by the use of similar instruments. These seismographs register on a revolving glass plate or drum automatically started by the earthquake motion, components of horizontal and vertical motions of the earth on a magnified scale, thus producing continuous diagrams, and indicating successive displacement of the ground in conjunction with the time. The account of the earthquake of December 28, 1885, the largest shock during the last three months, is here given as a sample of seismic record now issued in this country. The motion is slowly commenced, not accompanied by quick tremors, as is usually the case. At the 14th second from the commencement a considerable E.-W. motion occurred; in another second the maximum movement appeared in the same direction, which was followed by smaller shocks during about one minute; and from thence the oscillations, gradually subsided. As usual, the particles of the ground did not move to and fro, but traced a curvilinear path, although the E.-W. components always remained greater than the S.-N. components. In all, over 130 shocks or complete waves were registered. I may add in respect to the above earthquake, and in general, that the vertical motions are always — in our experience — smaller than horizontal ones, and the maxima and minima of these two kinds of motions are not synchronous. I shall have occasion before long to communicate to you the general results of all observations made during past years.

SEIKEI SEKIYA

The Imperial University of Tokio, Japan.

From *Nature* **33**, 603; 29 April 1886.

Mathematics

Hermann Grassmann was right

from Ian Stewart

THE late nineteenth century witnessed many attempts to develop an algebra of n -dimensional space, by analogy with the representation of the plane by complex numbers. Prominent among them was Hermann Grassmann's *Die lineale Ausdehnungslehre (The Calculus of Extension)*, published in 1844. Grassmann had the misfortune to write in a discursive, philosophical and obscure style at the time when axiomatic presentation was becoming *de rigueur* in the mathematical world. Further, he was a high-school teacher with no formal training in mathematics. Not surprisingly, his ideas were at first not taken very seriously — but later, Élie Cartan used Grassmann's 'exterior algebra' as the basis of his theory of differential forms. This work proved highly influential and established the importance of Grassmann's work. But, as Marielena Barnabei, Andrea Brini and Gian-Carlo Rota write in a recent paper (*J. Algebra* **96**, 120; 1985) "In the process Cartan paid a price of omission from which we are yet to recover." They argue that mathematicians have still not fully understood what Grassmann was trying to say, and propose a formulation of his ideas in modern terms which provides a natural language for the study of invariants.

In Grassmann's original work the same notation was used to denote two different operations performed on vectors, called the progressive and regressive products. This was not an oversight, but a deliberate choice. Every mathematician knows that what Bourbaki calls "abuse of notation" can, when wielded with care, greatly illuminate what would otherwise be a confusing situation. Unfortunately Grassmann managed to confuse greatly what might otherwise have been an illuminating situation. The tale has two main threads: differential forms and invariants.

Cartan's theory of differential forms attempts to make sense of the idea that expressions such as dx and dy in calculus can have a perfectly sensible meaning. The intuitive idea of 'a very tiny bit of x ' runs into severe difficulties, and is not very helpful; but the basic idea that the derivative dy/dx satisfies an equation $dy = (dy/dx)dx$ is crucial. So is the idea that an integral $\int dx$ can be thought of as an operation $\int$ applied to a differential form dx . The fundamental observation is that products of differential forms, such as $dx \cdot dy$, do not obey the usual laws of algebra. This is because the double integral $\iint dx \cdot dy$ represents a signed area, positive or negative; and interchanging x and y turns a left-handed coordinate system into

a right-handed one, multiplying all areas by -1 . In other words, one must set up the formation to make $dx \cdot dy = -dy \cdot dx$. Grassmann's progressive product obeys the same kind of law. Cartan found that the regressive product added nothing to his theory, and he quietly abandoned it. Until now, its demise has gone unmourned.

The new work of Barnabei *et al.* is part of an extensive programme of research that Rota has carried out for more than a decade, applying combinatorial methods to invariant theory. Classical invariant theory is the study of expressions that remain unchanged when their variables are transformed in particular ways. A simple example arises in connection with quadratic equations $ax^2 + bx + c = 0$, for which there is a well-known formula involving square roots of $b^2 - 4ac$. This quantity is called the discriminant, because it vanishes precisely when the two roots of the equation are equal. Suppose the variable is changed from x to $x+h$. Then the quadratic becomes $(x+h)^2 + b(x+h) + c$, or $ax^2 + (2ah+b)x + (ah^2+bh+c)$. Write this as $Ax^2 + Bx + C$. Now compute the new discriminant: $B^2 - 4AC = (2ah+b)^2 - 4a(ah^2+bh+c) = b^2 - 4ac$. Thus the formula for the discriminant does not change, although the variable x does. It is called an invariant of the transformation.

Invariant theory was invented around 1850 by Arthur Cayley and James Joseph Sylvester. It considers more complicated expressions than only quadratics, and more complicated transformations than $x \rightarrow x+h$. The fundamental problem in the theory is to prove that, in any particular case, the full system of invariants can be expressed in the terms of a finite subsystem, called a basis. The main classical method is to describe explicitly all invariants and to exhibit such a basis. By such methods Paul Gordan, the 'king of the invariants', proved in 1868 that for any expression in two variables, a finite basis exists. Limited extensions were proved by Gordan and others, but progress became painfully slow as ideas were overwhelmed by a swelling mountain of calculation.

In 1888 David Hilbert created a tremendous stir by proving that a finite basis always exists — and he did it without calculating any invariants at all. He used only very general abstract properties of systems of polynomials. "This is not mathematics", Gordan complained, "it is theology". But theology won the day, and explicit computation of invariants went out of fashion, being replaced by a much more abstract, higher level approach that

gave conceptual ideas pride of place. But history tends to run in spirals; and in recent decades there has been something of a classical revival, inspired in part by the needs of other branches of mathematics, and by applications. Explicit formulas, rather than existence results, are often desirable. Rota, an expert in combinatorics, has developed a powerful approach to such problems: his methods manage to be both conceptual and computational at once. How does one achieve such a balancing act? By making the notation reflect the concepts. As a result Rota has become the Edgar Allan Poe of mathematics, exhuming Grassmann's regressive product from its premature burial.

Barnabei *et al.* motivate their ideas in terms of three important types of transformation: orthogonal transformations are essentially rotations in a multidimensional space; symplectic transformations arise from mechanics; and a general linear transformation is, as its name implies, the most general transformation that keeps straight lines straight. Barnabei *et al.* summarize the standard results in the first two cases as follows: (1) a Hilbert space is a vector space having a symmetric form (x,y) . That is, $(x,y) = (y,x)$. A transformation T is orthogonal if it preserves this form, that is, $(Tx, Ty) = (x,y)$. Invariants under orthogonal transformations are polynomials in the symmetric form. Computations in Hilbert space are carried out in a naturally associated algebra, the Clifford algebra; (2) a symplectic space is a vector space having an antisymmetric (or symplectic) form $[x,y]$, that is, one such that $[y,x] = -[x,y]$. A transformation is symplectic if it preserves this form. Invariants under symplectic transformations are polynomials in the symplectic form. Computations occur in a naturally associated algebra, the Weyl algebra.

The parallel between these two theories is striking, and it suggests a fruitful question: is there an analogous setting for the general linear group? Barnabei *et al.* show that the answer is yes. They define a Peano space to be a vector space having a bracket $[x,y,z,\dots]$. Invariants under the general linear group are polynomials in the bracket. Computations are carried out in a naturally associated algebra, the double algebra. It is so called because it has two operations (rather than the usual one). They are, of course, modern versions of Grassmann's regressive and progressive product. *Plus ça change...*

Ideas that are ahead of their time do not always find immediate favour, and those that are not appreciated tend to be forgotten. But it is hard to keep a really good idea down. It was Grassmann's tragedy to be ahead of his time, but it is a mark of his genius that his ideas are now coming to full flower. □

Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Cell biology

Origins of the cytoskeleton

from Carl M. Cohen

A REPORT by Catherine Woods and Elias Lazarides on page 85 of this issue¹ shows that our understanding of membrane skeletal biogenesis shares at least one characteristic with the cells in which it has most recently been studied — both are embryonic. Lazarides and co-workers have recently provided us with fascinating insights into membrane skeletal biogenesis in chick embryo erythroid cells. Although these cells are nucleated, they nonetheless contain a membrane skeleton composed of the protein complex spectrin-actin-band 4.1 as well as the spectrin-binding protein ankyrin (goblin) which links spectrin to the integral membrane protein band 3. Lazarides' group has interpreted their observations in terms of a 'receptor-driven' membrane skeletal assembly, that is, that the amount of band 3 (the 'ultimate' membrane skeletal receptor) limits the amount of newly synthesized ankyrin that gets incorporated into the skeleton. Ankyrin is then thought to limit the amount of spectrin β chain (containing the ankyrin-binding site) which in turn limits spectrin α -chain incorporation². According to this scheme, spectrin chains that are not incorporated into the skeleton as α/β heterodimers, α/β_2 heterotetramers or higher hetero-oligomers are degraded^{3,4}.

A major question raised by this model is how the unassembled chains are targeted for destruction. Woods and Lazarides' new work¹ suggests that α and β chains not incorporated into the skeleton pass their brief existence as homo-oligomers, either α_2 or β_2 . These oligomers are largely incapable of forming stable functional associations within the membrane skeleton and the cytosol, and are somehow marked for destruction.

The possible existence of spectrin homo-oligomers runs counter to a small

body of literature suggesting that such complexes do not form, or are rare, in solutions of purified α or β chains⁵⁻⁸. But these studies were done on mammalian erythrocyte or brain spectrins, and there may be important differences between chicken and mammalian spectrins. Although chicken non-erythroid and human erythroid spectrin chains have much in common, these two forms are probably the most widely divergent among the spectrin-chain family⁹⁻¹¹. (Because chickens apparently possess only a single spectrin gene, the same must hold for chicken erythroid spectrin.) More to the point, recent unpublished studies by M. Mooseker, J. Coleman and J. Morrow at Yale University show that purified chicken erythroid β -spectrin chains have a remarkable propensity to self-aggregate. This is apparently true of fodrin chains as well¹². Taken together, these studies suggest that for most spectrins, self-oligomerization of spectrin chains, particularly the β chain, may be the rule rather than the exception.

Intriguingly, the chains making up homo-oligomers in the chick cytosol seem to be perfectly capable of forming heterodimers if the homo-oligomers are first denatured and then mixed with their partners under appropriate renaturing conditions. Thus, there is nothing inherently different about those chains that end up as homo-oligomers. What, then determines the proportion of spectrin chains that end up as heterodimers? Is the

proportion regulated by membrane-bound ankyrin? If not, what happens to the heterodimers that are synthesized in excess of ankyrin? Woods and Lazarides find no evidence for unassembled heterodimers in the cytosol of chick erythroid cells after translation *in vivo*. Are unassembled heterodimers also somehow targeted for destruction?

It seems likely that heterodimerization would take place in the cytosol during or soon after synthesis in order for the chains to be protected from self-oligomerization. In fact, the proposition that the α/β heterodimer (or α_2/β_2 heterotetramer) may form before its incorporation in the skeleton makes sense in the context of the known functional capabilities of spectrin chains. Several groups have shown that the high affinity association of spectrin in the spectrin-actin-band 4.1 complex requires an α/β heterodimer (or α_2/β_2 heterotetramer) even though isolated α and β chains bind to band 4.1 quite well⁸. Even fodrin, which can apparently bind to actin filaments without the need for band 4.1, must be heterodimeric for the association to occur⁷. Thus, it is more appealing to envision a fully competent heterodimer or heterotetramer forming a stable link in the nascent membrane skeleton than a single β chain arriving and waiting for its α mate.

If the homo-oligomer degradation route is operative in mammalian cells it may be less stringent, but the same principles may still apply. For example, in mice with any one of several genetic defects at the *sph* locus there is no α spectrin synthesized, whereas β spectrin is synthesized at a nearly normal rate. Almost normal amounts of β spectrin are bound to the reticulocyte membrane skeleton, and about 20 per cent of normal

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

More stamps for comet Halley (courtesy of Crown Agents Stamp company). First results from the Soviet, Japanese and European spacecraft encounters with comet Halley will be published in a special supplement in *Nature* on 15 May.

1. Lazarides, E. & Woods, C.M. *Nature* **321**, 85 (1986).
2. Lazarides, E. & Moon, R.T. *Cell* **37**, 354 (1984).
3. Blikstad, I., Nelson, W.J., Moon, R.T. & Lazarides, E. *Cell* **32**, 1081 (1983).
4. Woods, C.M. & Lazarides, E. *Cell* **40**, 959 (1985).
5. Calvert, R., Bennett, P. & Gratzer, W. *Eur. J. Biochem.* **107**, 355 (1980).
6. Yoshino, H. & Marchesi, V.T. *J. biol. Chem.* **259**, 4496 (1984).
7. Davis, J. & Bennett, V.J. *biol. Chem.* **258**, 7757 (1983).
8. Cohen, C.M. & Langley, R.C. *Biochemistry* **23**, 4488 (1984).
9. Birkenmeyer, C.S. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5671 (1985).
10. Glenney, J.R. & Glenney, P. *Eur. J. Biochem.* **144**, 529 (1984).
11. Wasenius, V.-M., Saraste, M., Knowles, J., Virtanen, I. & Lehto, V.-P. *EMBO J.* **4**, 1425 (1985).
12. Glenney, J.R. & Weber, K. *Analyt. Biochem.* **150**, 364 (1985).
13. Barker, J.E., Bodine, D.M. & Birkenheimer, C.S. in *Membrane Skeletons and Cytoskeletal Membrane Associations* (eds Cohen, C.M., Bennett, V., Lux, S. & Palek, J. (Liss, New York, in the press).

is retained in red cell membranes at steady state¹³. This suggests that whatever form the β chain is in, most of it makes it to the reticulocyte membrane skeleton without being degraded. Subsequently, 80 per cent of the β chains are degraded, presumably because they fall off the membrane (as oligomers?). These and other instances documented in mice¹³ show that the mere presence of single chains (be they monomers or homo-oligomers) does not dictate immediate destruction. Possibly, if the chains make it to the membrane soon enough they are spared. In cases where there are no heterodimers to compete for membrane binding, the β chains may have a higher probability of associating with the membrane and escaping degradation, if only transiently.

Interesting as these studies are for those

in the membrane skeleton field, there are also paradigms emerging which may be more general. The notion that post-translational folding, and hence the fate of a protein, may depend as much on the presence or availability of other protein subunits as on primary structure clearly has wide implications (see ref. 2 for review). Also, the rather unorthodox characteristics of the degradation of excess α and β chains (the former ATP dependent and lysosome-like, and the latter ATP independent) suggests that more surprises are in store the deeper we probe membrane skeletal beginnings. □

Carl M. Cohen is in the Department of Biomedical Research at St Elizabeth's Hospital, Tufts University School of Medicine, Boston, Massachusetts 02135, USA.

Information technology

New flow of ideas in networks

from Alistair Mees

Flows in networks are a common feature of today's world — they may involve physical quantities such as traffic in a city, or information in a telephone or computer network. Many problems that do not at first sight directly concern network flows can easily be converted to that form: railway scheduling and project management are well-known examples. Until recently only simple versions of such problems have been easy to solve. But now a new mathematical approach opens the way for greatly improving the efficiency of the solution of common, hard problems.

With the advent of ever more complicated networks, particularly in telecommunications, rapidly increasing amounts of computer time are needed to answer questions such as: what is the capacity of a particular network and how should the flows be arranged to attain it; and what is the cheapest way to expand a network to have a particular capacity?

For the simplest sort of network flow, in which there is only one commodity flowing (for example a network of pipes carrying water, or an electricity grid), many problems can be solved by existing methods (Ford, L.K. & Fulkerson, D.K. *Flows in Networks* Princeton University Press, 1962). The methods are good both because they are intuitively 'right' — a feature which is more important in mathematics than is often admitted — and because they are efficient for use on computers. For instance, think of a network with capacity limits on the arcs (the lines), where there is a single source and a single destination, and we want to find the maximum traffic capacity. The famous Ford–Fulkerson method finds a route from source to destination; looks for the bottle-

neck on it which determines the maximum capacity of that route; and pushes through that amount of traffic. That particular route is now blocked because the bottleneck is operating at maximum capacity. The method now finds another route and does the same, continuing until all routes are blocked by previous traffic. We now have the maximum flow that the network is capable of carrying.

More sophisticated problems, for example those in which there are costs associated with different routes, can be solved by related methods. The 'out-of-kilter' algorithm is one, in which certain conditions are checked for each arc, which it is said to be out of kilter if they do not hold: the algorithm uses a route-search-and-block method successively to put arcs into kilter. (This is the only time I have ever heard the expression 'in kilter'.)

The most important mathematical feature of these methods is that they are based on results from duality theory. Duality refers to the fact that for every optimization problem there is a related problem which (under some convexity conditions that hold in the cases under discussion) has the same solution. For example, the basic Ford–Fulkerson result is actually a duality theorem which states that maximum flow equals minimum cut, meaning that if the network is cut so that the source is in one part and the destination in the other, the sum of the capacities of the broken arcs is always at least as big as any allowed flow, and for certain cuts it will equal the biggest possible flow. Duality theory pervades the mathematical theory of optimization and often has a price or a game-theoretic interpretation: if we let the capacity of an arc be the cost of

breaking it, the Ford–Fulkerson result could be expressed in terms of an attacker who tries to isolate the source from the destination, and a defender who tries to make the attack as expensive as possible. The conditions for an arc to be in kilter in the more general problem come directly from duality theory.

Things get harder when we talk about multicommodity flows, in which there may be distinct kinds of traffic, with different sources and destinations, all sharing the same capacitated network. This is a lot closer to many real-world problems, but the ideas that were good for single-commodity flows do not work well: if one kind of traffic is maximized, another kind may be blocked, although some different arrangement would allow both flows to coexist. Trying out different arrangements from scratch quickly leads to a combinatorial explosion in which the number of options to try grows exponentially fast with the size of the network or with the number of commodities. Exponential growth of the time taken to solve a problem means that beyond a certain size of problem (in this case probably rather a small size) it becomes impossible to find solutions with present technology, and any future improvements in computing power will only marginally increase the size that can be managed. There is nothing for it but to look for better algorithms.

Multicommodity problems have been solved by linear programming methods (Assad, A.A. *Networks* 8, 37; 1978), but because they transform to very special and degenerate linear programs this is an inefficient way to proceed: only problems of a rather modest size can be solved.

Recent work by Soviet mathematicians shows promise in solving multicommodity problems in the same sort of nice way that single-commodity problems are solved. An entire issue of the journal *Discrete Applied Mathematics* (11; 1985) is devoted to a paper by M.V. Lomonosov describing combinatorial approaches to multicommodity flow problems. There are many theoretical and algorithmic ideas in this long paper, some of them new and some of them summarizing recent work by the author and his colleagues, particularly A.V. Karzanov, most of which has not previously been published outside the Soviet Union. Duality theory is as central as ever, with Ford and Fulkerson's max-flow min-cut theorem being extended, but perhaps the most practically useful ideas are switching techniques that give systematic ways to unlock a multicommodity network and allow different flow patterns. Switching and draining transformations are introduced to allow flows to be rearranged without reducing the traffic amounts. Providing systematic transformations between different flow patterns is an important step towards overcoming the combinatorial explosion; the

fact that these transformations do not reduce flow may mean that, as with the simplex algorithm for linear programming, rather a small number of different arrangements can be stepped through to reach the optimum.

It is too early to say whether these new ideas will lead to significantly greater efficiency in solving multicommodity problems. The paper deals only with certain specialized flow problems, so it is roughly analogous to the situation with the Ford-Fulkerson approach before the out-of-

kilter algorithm was developed. Lomonosov does not discuss algorithmic efficiency directly, nor does he give explicit examples, but it is likely that his methods will prove good for some kinds of problems. If this turns out to be the case, we may hope for extensions to more practically realistic problems and perhaps a large increase in the size of multicommodity problems that can be solved. □

Alistair Mees is in the Department of Mathematics at the University of Western Australia, Nedlands, Western Australia 6009, Australia.

Molecular biology

Self-incompatibility in plants

from John Fincham

In many families of flowering plants, exemplified by the Solanaceae (which includes the genera *Nicotiana* and *Lycopersicon*), self-pollination is prevented by a genetic system based on multiple alleles at what appears to be a single locus which is given the symbol *S*. Impressive progress towards the goal of discovering what the products of the *S* allele are and how they relate to each other in molecular terms has recently been made by two groups, first by J. B. Nasrallah *et al.* from Cornell University (*Nature* **318**, 263; 1985) and now by Anderson *et al.* of Melbourne University and reported elsewhere in this issue (*Nature* **321**, 38; 1986).

It is known that pollination is blocked only when the same allele is present in the germinating pollen grain and the style down which it is trying to grow. Because wild populations of outbreeding species usually have tens or even hundreds of *S* alleles, this system ensures that the great majority of random cross-pollinations are fertile. In some plant families, including the Cruciferae and its best-known genus *Brassica*, there is a variant system in which pollen expresses one or both of the *S* alleles present in the diploid plant which bore it, even though each pollen grain, as a haploid product of meiosis, retains only one allele. The difference between the two systems is presumably one of timing of *S*-locus expression — post-meiotic in the Solanaceae and pre-meiotic in the Cruciferae — and does not necessarily imply any fundamental difference in the functions of the products of the respective *S* loci. In each case, incompatibility evidently results from some kind of interaction between products of identical *S* alleles.

What is the mechanism of this self-recognition by *S* alleles? There are two possibilities. If one supposes that the *S* locus is really a single gene, encoding the same protein in pollen and style, one can think in terms of dimerization of similar protein monomers, located respectively on the cell surface of the pollen tube and in

the cell sap of the style. The alternative is to postulate that the *S* locus is a complex of at least two genes expressed respectively in pollen and style, with each *S* allele encoding a stylar product and a different but complementary pollen product. The second model does not require the products of any particular *S* allele (more accurately termed a haplotype) to bear any specific structural relationship to those of any other — each haplotype could evolve independently, subject to selection only for self-recognition. A bipartite structure of the *S* locus has, in fact, long been postulated as an explanation of *S* mutations, which tend to affect pollen and style reactivities independently.

The breakthrough by the Cornell group was the molecular cloning of a complementary (c)DNA sequence representing an abundant style-specific messenger (m)RNA in *Brassica oleracea* of genotype *S₆S₆* (obtained by forced self-pollination). When used to probe restriction-endonuclease digests of DNA from plants carrying other *S* alleles, this cDNA clone revealed a restriction fragment size polymorphism that was closely and perhaps inseparably linked to the *S* locus; each *S* allele was associated with its own specific array of restriction sites within the region covered by the probe. Moreover the cDNA, when cloned into an expression vector, was translated in *Escherichia coli* into a protein product that reacted with an antiserum made against a glycoprotein of relative molecular mass (*M_r*) ~ 85,000 that had been found specifically in styles of plants bearing allele *S₆*. Only a partial sequence of the cDNA clone has been published so far.

The starting point for the Australian group was the finding of an abundant glycoprotein of *M_r* ~ 32,000 in styles of *Nicotiana glauca* (tobacco) plants carrying allele *S₂*. Somewhat surprisingly, no glycoprotein of comparable abundance was isolated by the same method from plants carrying other *S* alleles, which raises the

possibility that different *S* alleles have very different stylar products. Determination of part of the N-terminal amino-acid sequence of the *M_r* 32,000 protein permitted the synthesis of a mixed DNA probe for the encoding gene, and the gene was indeed recovered by this means from a cDNA library made from mRNA of styles carrying *S₂*. Sequencing of the cDNA clone revealed an open reading frame encoding a 204-residue polypeptide chain with a 22-residue N-terminal hydrophobic portion (presumed signal sequence) leading into the sequence already found at the N-terminus of the mature glycoprotein. The cDNA hybridized strongly to DNA of plants carrying *S₂* and much more weakly to DNA of plants with other *S* alleles, suggesting again that the different *S* alleles have very substantially different coding sequences and only remotely related stylar products. There is as yet no direct evidence about the extent and nature of the inter-allele differences within *N. glauca*, but Anderson *et al.* do present some preliminary information from another Solanaceous species, the wild tomato *Lycopersicon peruvianum*. A part of the N-terminal sequence has been obtained from two different stylar glycoproteins associated with different *Lycopersicon* *S* alleles; these are clearly homologous with each other (although there are also several differences) and also with the N-terminal sequence of the *N. glauca* *S₆* protein. Disappointingly, but not surprisingly, no similarity is visible as yet between the *Nicotiana* or *Lycopersicon* sequences and the limited amount of sequence available from the *Brassica* *S₂* protein.

Clearly we are at the beginning of what is bound to be a period of intense activity as these highly promising observations are carried further. Several questions are now readily answerable. We should soon see some full-length comparisons between the DNA sequences of different *S* alleles (or at least those parts of them governing the stylar products); it should soon become apparent if there is any homology between the *Nicotiana/Lycopersicon* and *Brassica* *S* loci. More difficult, but probably not long delayed, will be the identification of *S*-allele products in the pollen tube and the cloning of the corresponding DNA sequences. If the *S* locus really does have separate pollen- and style-specific products with separate coding sequences, much may depend on how close together these sequences are in the genomic DNA. If they are separated by only tens of kilobases, walking from one to the other should not be difficult; but if they are hundreds of kilobases apart, as they could be and still be almost inseparably linked, there will be more of a problem. □

John Fincham is Professor of Genetics at the University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

Keeping supercomputers under control

SIR—The Purdue Water Resources Research Center in conjunction with the Office of Health and Environmental Research of the US Department of Energy recently co-sponsored a meeting on the current uses of, and future needs for, supercomputers in hydrology. 'Supercomputers in Hydrology: Future Directions', the first meeting of its kind, aimed to evaluate the need for large-scale vector and parallel processing machines and their peripherals in the hydrologic sciences, especially in subsurface hydraulics. While the charge of the seminar was to specifically deal with hydrologic applications of supercomputers, the conclusions drawn at the meeting apply equally well to many other fields of science and technology.

Almost all technical issues involving the use of supercomputers were presented either by a computer scientist, a programmer or an applied mathematician. Because the architecture of most supercomputers is radically different from scalar processing machines, this is no surprise. The real importance of this observation is the implied increase in need for multidisciplinary research teams to solve major ecological problems. Unfortunately, in the United States, many of the larger supercomputer funding initiatives do not address this issue at all.

This defect in the supercomputer initiatives leads to two major problems: scientists not doing 'science' and the inefficient use of supercomputers. If funds were available to support programmers with expertise on a machine of a certain architecture, the scientist would only need to communicate effectively with the programmer. The projected changes in architecture affect the programmer and not the scientist; hence the scientist is free to do science.

An alternative to long-run funding of programmer-consultants is for the initiatives to increase their effort in the development of very high level languages. Rice of Purdue's Computer Science Department gave a talk on projected advances in hardware and software over the next 10 years. From 1975 to 1985 computing speed increased by a factor of 25. Between 1985 and 1995 the speed is projected to increase by a factor of 2,000. On the other hand, from 1975 to 1985, computer languages have been improved by a factor of only 1.4. Unless more funding is made available for the development of high-level languages, improvement in ease of coding is not expected to increase nearly as fast as machine computational power. In fact, because of radical changes in architecture, programming may well become more difficult.

Another important point brought out at

the seminar is that advances in science should not be overpowered by advances in computational capability. Do we really improve our fundamental understanding of science by looking at larger and larger computational problems? Many would argue that the understanding of complex chemical and physical interactions in the subsurface and the ability to estimate parameters is so poor that the large supercomputer models may in fact be totally incorrect. In this instance, supercomputer modelling efforts may turn out to be a greater obstacle to real progress than no model at all.

So there is clearly a role for supercomputers, but the extent of this role will depend on the availability of high level languages, expert programmer consultants working closely with multidisciplinary teams, exposure to different computational systems, and a judicious choice of problems with the aim of advancing our understanding of science.

JOHN H. CUSHMAN

*Water Resources Research Center,
Lilly Hall of Life Sciences,
Purdue University,
West Lafayette,
Indiana 47907, USA*

FRANK J. WOBBER

*US Department of Energy,
Office of Energy Research,
Washington, DC 20545, USA*

The real hazards of nuclear fallout

SIR—In his article "On minimizing the consequences of nuclear war"¹, Carl Sagan is misleading in his discussion of the dangers of radioactive fallout. Our remarks on this question are made in the spirit that the goal of scientists is to be as accurate and unbiased as possible in assessing the potential or alleged impacts of an event such as nuclear war, and that either minimizing or maximizing these consequences is a disservice to the general public and to decision makers who seek guidance. If scientists are not accurate, credibility is lost and messages from the scientific community will eventually be ignored.

To our knowledge, the serious consequences of local (or early) fallout, that which is deposited during the first 24 hours, have been accepted by the scientific community for over three decades. Projections of the intensity and extent of local fallout are highly sensitive to a number of variables, which helps explain why different assessments have produced widely different results. Uncertainties in these projections can be divided into three categories, those due to (1) the targeting scenario, (2) the fallout calculation model and (3) the selected meteorological conditions.

The targeting scenario contains variables such as numbers of weapons, the yield mix, fission fraction, height of burst and precise target locations. The height of burst (HOB) is of particular significance because air bursts do not produce significant local fallout, except for the possible rainout of debris from tactical yield weapons. Only when the fireball interacts with the ground (a ground level or near ground burst) does significant local fallout ensue. A reasonable assumption often made is that hardened military targets are targeted with ground bursts. For the 'softer' industrial and other military targets, maximum damage is accomplished by air bursts where the HOB can be optimized. The primary fires hypothesized in urban areas in 'nuclear winter' studies are assumed to be initiated by airbursts, because ground bursts have shorter range for initiating fires. Hence, maximizing smoke production implies minimizing local fallout. (2) Uncertainties in dose calculations in the best fallout models originate from several sources. These uncertainties are due to limited experimental calibration data, whether the modelled radioactivity is rigorously conserved, and whether time of arrival is properly accounted for. (3) Assumptions about selected meteorology — such as wind velocities, shears, precipitation patterns — affect the results. Hence local fallout assessments can vary greatly depending on these many assumptions.

In assessing the human impact of local fallout, additional factors must be considered. By far the most sensitive of these is the protection factor afforded by structures, such as homes, buildings, basements, and shelters. These structures can dramatically mitigate the unprotected dose assessments normally cited. An additional important consideration is the assumed lethal acute external whole body dose levels. Finally, for radiation exposure that is protracted in time, biological repair of the resulting damage is important in mitigating the effects². This is especially important with regard to global fallout, where the dose is received slowly over many years. Dose effectiveness factors from 0.1 to 0.5 for chronic exposures have been suggested³. This means that a large chronic dose will have an effect equivalent to a much smaller acute dose. Any presentation that implies that our planet would be a radioactive desert of certain demise is not including these important factors in a balanced sense; hence, inaccurate and biased estimates can be created. Calculations of total fatalities produced by large-scale attacks on the continental United States undertaken by our group have produced estimates of fallout fatalities (after subtracting those already killed by blast and thermal effects) that range over almost two orders of mag-

nitide. This large variation in fallout fatalities is well understood in terms of variations in the parameters discussed above.

As part of our contribution to the recent International Council of Scientific Unions (ICSU) SCOPE-ENUWAR study of the environmental effects of nuclear war³, we undertook calculations of local fallout using the computer model KDFOC2⁴, using single burst fallout patterns and a semi-empirical approach for multiple explosions. Calculations were done for a 6,000-megaton multi-phase counterforce-countervalue nuclear exchange. Results indicated that about 7 per cent of the land areas of the United States, Soviet Union and Europe would receive a minimum of 450 rads within 48 hours. This assumed no shielding and would apply to a population unprotected by structures. The SCOPE-ENUWAR study³ was primarily concerned with direct effects on unprotected plants and animals and excluded direct effects on humans. In considering the direct effects on humans, protection factors and population distribution should be added into the picture, with the result that land area envelopes are reduced below 7 per cent by factors of several (land area envelopes scale very roughly as the inverse of the protection factor). As an example, a recent study⁵ used protection factors of 3 for rural areas and 10 for urban areas. The figure of 30 per cent of all northern-midlatitude land areas receiving an acute mean lethal dose to humans obtained by TTAPS and quoted by Sagan¹ is clearly much higher than our SCOPE result, and we believe inaccurate and misleading.

There is a hint of *On the Beach* in Sagan's discussions of global fallout; the spectre of a cloud of lethal levels of radioactivity spreading around the globe. Our group first reported in 1983 that long-term total doses of 100 rem or larger in global fallout hot spots are possible⁶. Sagan refers to this, and then interjects the possible effects of a 100 rem acute dose on the immune system (a speculative new result). This may mislead the reader as there is an immense difference in the effects of an acute compared with a long-term (lifetime or 50-year) dose.

The calculations at Livermore on global fallout reported in 1983⁶ were supplemented by additional work in 1984⁷. The original results using the GLODEP2 code⁸ assumed an atmosphere unperturbed by smoke. These were supplemented by new work utilizing GRANTOUR⁹, a fully interactive microphysical, three-dimensional global circulation model code developed at LLNL. Doses were calculated with and without smoke in the atmosphere to assess the impact of a perturbed climate on global fallout.

Our GLODEP2 calculations for strategic nuclear exchanges of 5,000 to 6,000

megatons predict that the 50-year unsheltered unweathered average external total body gamma dose levels in the Northern Hemisphere would be about 15 rads, and about 0.5 rad in the Southern Hemisphere. The maximum of 30 to 40 rads appears in the 30° to 50° north latitude band. Values predicted for the global population (chronic) dose are typically about 6×10^{10} person-rads. The dose in hot spots, obtained using an area grid of 10° latitude by 10° longitude, are a factor of 6 to 8 higher than the Northern Hemisphere averages. These results have an estimated confidence level of about a factor of 2. Some 50–75 per cent of the global fallout dose is due to the tropospheric injections of radionuclides that are deposited in the first month.

The additional calculations, utilizing GRANTOUR involving a perturbed nuclear winter atmosphere indicate that the above dose assessments would be about 15 per cent lower in the Northern Hemisphere and marginally higher (approximately 1 rad) in the Southern Hemisphere compared to the unperturbed atmosphere predictions. These results are consistent with the projection that smoke injections can increase vertical stability, inhibit precipitation, and increase interhemispheric transport. We note that the GRANTOUR calculations utilizing a fully interactive microphysical approach with a three dimensional global circulation model produced results that were in good general agreement with the GLODEP2 results that depended on empirical observations from past atmospheric tests.

These calculations have been presented at a number of scientific meetings, including the ICSU-SCOPE workshop on radiation held in Paris in October 1984. There, an international array of radiation experts reviewed this work in detail, and it then became a basis of the chapter of radioactivity in the SCOPE-ENUWAR report³ and underwent additional external review by dozens of international experts. The effects of these levels of global fallout were summarized in the Report of the Paris Commission on Radiological Dose Assessments and Biological Effects¹⁰. It concluded "The long-term increase in genetic and carcinogenic effects on humans from global fallout is of the order of 1 per cent of the natural incidence and should be considered a second order effect." No mention is made of pro-dromal effects on humans because at these lifetime dose levels and considering biological repair, pro-dromal effects from global fallout would not be observed. This does not sound anything like the *On the Beach* scenario or the impression transmitted to the audience at the Washington Sheraton Hotel briefing held by Sagan and others on Halloween, October 1983.

It is true that our original calculations have not been published in a refereed

journal, nevertheless they have undergone extensive external review. Sagan should perhaps heed the Biblical injunction "let he who has not sinned cast the first stone". Reference 2 in his paper¹ (EOS, 63, 1018; 1982) is a 250-word unreviewed abstract of a paper that was to be given at the San Francisco meeting of the American Geophysical Union but was cancelled before presentation. Authored by the "TTAPS team", this unpublished paper was cited in Sagan's recent *Nature* article 10 times!

This work was performed under the auspices of the US Department of Energy by the Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

JOSEPH B. KNOX

*Atmospheric and Geophysical Science Division,
Lawrence Livermore National Laboratory,
Livermore,
California 94550, USA*

CHARLES S. SHAPIRO

*Department of Physics,
San Francisco State University,
1600 Holloway Avenue,
San Francisco,
California 94132, USA*

1. Sagan, C. *Nature* **317**, 485–488 (1985).
2. NCRP84 National Council on Radiation Protection and Measurement, Report 64 (Government Printing Office, Washington, DC, 1980).
3. Pittock, A.B., Ackerman, T.A., Crutzen, P., MacCracken, M., Shapiro, C.S. & Turco, R.P. *The Environmental Consequences of Nuclear War Vol. 1* (Wiley, Chichester, 1985).
4. Harvey, T.F. & Serduke, F.J.D. *Fallout Model For System Studies* (Lawrence Livermore National Laboratory Report UCRL-52858, 1982).
5. Daugherty, W., Levi, B. & von Hippel, F. *Symposium on Medical Aspect of Nuclear War*, Washington DC, (National Academy Press, Washington, 1986).
6. Knox, J.B. *Proceedings of the Third International Conference on Nuclear War*, Erice, Sicily, Italy (Laboratori Nazionali di Frascati dell' INFN); Lawrence Livermore National Laboratory Report (UCRL-89907).
7. Shapiro, C.S. & Edwards, L.L. *Scenario and Parameter Studies on Global Deposition of Radioactivity Using the Computer Model GLODEP2* (Lawrence Livermore National Laboratory Report UCID-20548, 1984).
8. Edwards, L.L., Harvey, T.F. & Peterson, K.R. *GLODEP2: A Computer Model for Estimating Gamma Dose Due to Worldwide Fallout of Radioactive Debris* (University of California, Lawrence Livermore National Laboratory Report UCID-20033, 1984).
9. Walton, J.J. & MacCracken, M.C. *Preliminary Report on The Global Transport Model GRANTOUR* (Lawrence Livermore National Laboratory Report UCID-19985, 1984).
10. Report of Commission of Radiological Dose Assessments and Biological Effects – Paris Workshop (SCOPE-ENUWAR Newsletter, October 1984).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a rather technical character which are not provoked by articles or by letters previously published (where the Matters Arising section remains appropriate).

Group selection revisited

Eric Charnov

Evolution Through Group Selection. By V.C. Wynne-Edwards.
Blackwell Scientific: 1986. Pp. 386. Hbk £29.50, \$58; pbk £14.95, \$29.

NEARLY 25 years ago, V.C. Wynne-Edwards published a provocative and important book about the evolution of social behaviour. It was entitled *Animal Dispersion in Relation to Social Behaviour*, and in it Wynne-Edwards proposed that most features of vertebrate reproductive and social behaviour are adaptations whereby populations regulate their numbers and so avoid over-exploitation of their food resources. Most of the book was a discussion of biological examples of this process at work. Thus, clutch size, territoriality, dispersal, male display and competition for females, as well as other aspects of social interaction, were assigned functions in terms of population regulation — individuals practised reproductive restraint when necessary for the good of the group's resource management. Of course, such group-level adaptations would seem to require that natural selection acts mainly on the group, and not through individual reproductive success.

Animal Dispersion had an immediate impact. It drew an enormous following, and during the 1960s various ideas based on "group selection" became popular among many biologists and social scientists. The book's appeal may have stemmed from the intellectual climate of the time, combined with the facts that the main thesis was both morally satisfying and a *tour de force* of radical thinking: to explain so much with one simple principle!

Wynne-Edwards himself realized that *some* form of group selection was required by his theory; he clearly said so. And his main critics (D. Lack, W.D. Hamilton, J. Maynard Smith and G.C. Williams) immediately fired back. To them it seemed unlikely that natural selection at a group (or population) level could overcome selection at the level of individuals. Furthermore, most of the biological examples he discussed, and which he saw as beneficial in regulating a population, could, they considered, be more easily explained in terms of individual reproductive success. This latter criticism was even more important than those of the notion of group selection; alternative hypotheses, not considered by Wynne-Edwards, did a better job with the data, and his vision of individual reproductive restraint was illusory. So said the critics. However, by taking such a strong stand on the "why's" of social evolution, his book focused attention on the subject: ideas that were developed over the past two decades in oppo-

sition to his theory are prominent in much of the current thinking in sociobiology and behavioural ecology.

This new book is an up-date of *Animal Dispersion*. The first half includes a re-statement of the hypothesis that populations manage their resources; again, most social interaction (at least in vertebrates) is seen as operating to regulate the population concerned. Following this is an

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Subject of study — the red grouse, well-camouflaged in its natural habitat.

account of the population and social dynamics of red grouse, the fruit of a superb long-term field study. In the second half of the book Wynne-Edwards discusses group selection as a process, both in theory and as seen against the data. He introduces one major new "population level benefit" hypothesis, describes a few case histories and ends with a good summarizing chapter. His new hypothesis, to which I shall come in more detail at the end of this review, is that populations benefit from managing their genetic structure; somehow, natural selection has favoured populations which control their mating system (for example monogamy versus polygamy), dispersal and other genetic characteristics which may be involved in the working of group selection. Simply, these controls exist and they function to facilitate the evolution of group benefits. They both facilitate group selection, and are themselves examples of it.

This new book is bound to be widely read and discussed. In my opinion, however, it suffers from many of the same problems as its predecessor. At issue here is not the *possibility* of group selection,

but whether it has produced all (or many of) the behaviours and reproductive phenomena that Wynne-Edwards considers.

To develop this point, let us first consider the "resource management" hypothesis. Here I will pose three questions. Is it needed (in theory)? Is it supported by the evidence? Is it a likely outcome of a group selection process? This critique applies equally to the conceptual development of the author's theory and his case history of red grouse.

Wynne-Edwards claims that predator populations *must* regulate their numbers socially or they would typically decimate their prey populations; this applies also to herbivores and their food plants. But I know of nothing in population dynamics theory which generally supports this proposition. Numerous equations describing predator-prey interaction have stable or cyclical solutions. Predators may often be limited by their own predators, by parasites or by the availability of food of adequate quality. Density-dependent birth and death effects may operate in many ways, and may often stabilize the population at numbers far short of those required for resource destruction. The fact that individuals use social interaction to compete for food, space or mates does not necessarily support any hypothesis of population regulation; indeed, even if social interaction *does* sometimes regulate population size, such a density-dependent effect may simply be viewed as a by-product of individual reproductive competition. Wynne-Edwards's blanket claim — that we need social regulation or else . . . — is much too strong. (While I accept his key observation that predators do not usually decimate their prey, I doubt that anyone has studied how often this does or does not occur.) Of course, that "resource management" may not be needed *generally* does not mean it does not or cannot ever happen.

In the past 20 years we have greatly increased our knowledge of social evolution and natural selection, including group selection. Wynne-Edwards discusses a fair bit of this newer work on group selection, but he mentions none of the rest of sociobiological thought. Modern ideas on mating systems, sexual selection, sex allocation, territoriality and group structure, game theory and evolutionarily stable strategies — almost the entire contents of books such as *Behavioural Ecology*, edited by J. Krebs and N. Davies (Blackwell Scientific, 1984) or M. Daly and M. Wilson's *Sex, Evolution and Behavior* (Duxbury, 1983) — are virtually ignored. This is perhaps the book's greatest failing. At the very least, the author ought to have attempted to convince us by a fair contest of various ideas set against the evidence, especially when many workable alternatives were put forth in response to his

Aquila

previous book. Although plenty of data are discussed, critical tests of hypotheses are not. We are, then, left with a work of advocacy.

Professor Wynne-Edwards appears to have adopted the position that the criticisms of *Animal Dispersion* simply mandated his finding, demanding only that he should demonstrate a mechanism of group selection which he has now found in the works of M.J. Wade and D.S. Wilson.

"Group selection is not different from natural selection. ..."

The concept of group selection has had a chequered history, and here I can give only a necessarily simplistic account of the arguments. Readers who wish to follow things up should refer to the review articles by M.J. Wade (*Q. Rev. Biol.* **53**, 101–114; 1978) and D.S. Wilson (*A. Rev. Ecol. Syst.* **14**, 159–187; 1983).

Group selection is *not* different from natural selection. Suppose a population is split into groups where the members within a group interact mostly with each other. As a result of social interaction, feeding, predator escape and so on, individuals produce offspring. Clearly, natural selection will favour attributes (genes building phenotypes) which confer higher reproductive success. But a phenotype's reproductive success or fitness may be thought to consist of two parts: the first is an individual's contribution of genes, relative to other members of its group; the second is the group's contribution of genes to the overall population. These are the within-group and between-group components of fitness. While not all theorists agree with this distinction, it is useful in showing that attributes may sometimes be selected for if they increase group productivity, even if their bearers do worse at reproducing relative to other members of their own group.

Of course, some attributes may increase *both* group productivity *and* an individual's fitness relative to other group members; these two components of fitness need not conflict. The strength of each component depends upon factors such as group size and stability, how groups are formed, dispersal patterns and so forth. Using a fairly extreme form of differential group productivity (extinction), Wade has demonstrated in laboratory experiments with *Tribolium* that the between-group component of fitness is sufficient to produce a fair amount of phenotypic change. A group-selection effect has also been shown in the field for the evolution of sex ratio in group-structured populations of parasitoid wasps and mites. Briefly, a sex ratio of 1:1 is favoured within each group, but the migrants or dispersers are all females and so the group productivity component of fitness favours a female-biased sex ratio. It is the within-group

component of fitness that favours an equal sex ratio. The resulting sex ratio is often female biased, reflecting a balance of the between- and within-group components of fitness. This group-selection interpretation for the female-biased sex ratio is not accepted by everyone, and other explanations are possible. Yet I think it is a reasonable interpretation. So, the between-group component of fitness can influence evolution; it can sometimes affect the value of characters. In his 1983 review, Wilson suggested that the evolution of traits containing a strong component of self-sacrifice would probably require small groups with one or a few founders (or kin selection, see below), and he made the sensible suggestion that the traits most likely to be favoured with more diffuse groups would be those with high social (group) benefit and little personal costs. Altogether, biologists would do well to pay closer attention to the thoughts of Wilson and Wade.

Wynne-Edwards is doing much more, however, than arguing that selection can sometimes favour traits which benefit the group. He is claiming that such traits will *usually* evolve and will include traits for population regulation for resource management. He further asserts that such traits will often not be opposed by the within-group component of fitness; what benefits the group will usually benefit the

"Wynne-Edwards is doing much more, however, than arguing that selection can sometimes favour traits which benefit the group. ..."

individual within the group. Although this latter assumption supports his hypothesis, I can think of no compelling reason why we would find it to be true for the multitude of examples that Wynne-Edwards discusses. Furthermore, nothing in the literature particularly suggests population regulation as the typical or expected outcome; the example of sex ratio in wasps and mites was not at all related to potential over-exploitation of food resources.

Another response to *Animal Dispersion* was the development of the theory of kin selection, mainly by W.D. Hamilton. There are several approaches to this question, one of which shows kin selection to be a special case of group selection, as discussed above. In the new book, Wynne-Edwards shows a curious love-hate relationship with kin selection. Early on he claims that it is a form of group selection; then, in the final two chapters, he totally rejects it in favour of group selection. Scattered throughout the book are various discussions of the inadequacy of kin selection to explain this or that; these are invariably followed by a glib claim that group selection will do the trick. I really do not know what to make of such statements.

I will end this review with a brief discussion of Wynne-Edwards's new proposal: that is, populations genetically structure themselves to foster beneficial processes such as group selection. I find this notion difficult to understand in terms of population-level adaptation. Let us consider an example. In Chapter 16, Wynne-Edwards uses the idea to explain a long-standing puzzle in the reproductive biology of Atlantic salmon. Immature fish (called parr) first spend a few years in fresh water, and then go to the sea to feed, grow and mature, before returning as adults to spawn in fresh water. The puzzle is that some parr (young males, who have not yet gone to sea) become reproductive, spawn and then turn off sperm production to join the usual out-migration to the sea. Some of these parr will undoubtedly return to spawn as adults. Wynne-Edwards's answer to this puzzle (why do parr become reproductive?) is that these fish are assigned the role of keeping genes in the group. Field data suggest that many adult fish do not return to their home stream to spawn; thus, the parr are the main bearers of in-group genes to the local group and this role is a form of genetic structuring favoured by group selection.

Three observations are in order. First, the population structure (with lots of mixing of fish from different groups during reproduction) is just that where group selection would not really work. Second, the population-level benefits of this "inbreeding" are not really specified (and totally escape me). Third, small breeding males are common in fish species and have now been studied in salmonids. The idea of the frequency-dependent fitness advantage of small males, in reproductive competition with larger males, has proved a very powerful tool in the analysis of these systems; this involves no postulated group benefit whatsoever. Readers of Wynne-Edwards's present book should compare carefully his explanation for the existence of these small breeding males with the evidence and interpretation of the same phenomenon by M. Gross (*Nature* **313**, 47–48; 1985). A fair contest of ideas would seem to be in order.

Animal Dispersion, or rather the responses to the extreme position adopted by Wynne-Edwards in that book, did much to discredit notions of group selection. For many people, the pendulum has swung from uncritical acceptance to equally uncritical rejection of the idea. I hope that the similarly extreme tone of *Evolution Through Group Selection* does not add further to that process. Natural selection, operating in a group-structured population, can sometimes produce the unexpected. And we still have so much to learn. □

Eric Charnov is a Professor in the Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA.

Coming in to the cold

Daniel Branton

Low Temperature Methods in Biological Electron Microscopy. By A.W. Robards and U.B. Sleytr. Elsevier:1985. Pp.551. Hbk \$111, Dfl. 300; pbk \$48.25, Dfl. 130.

GREATER USE of low-temperature methods and frozen specimens in biological electron microscopy has resulted from recent advances in a variety of morphological and analytical applications, especially freeze-etching, immunocytochemistry and the direct examination of hydrated frozen specimens. While all of these applications depend upon the use of frozen specimens, workers experienced in one particular area will, upon reading Robards and Sleytr's book, discover well-developed traditions and critical manipulations that have simply not crossed the boundary between, for example, those laboratories engaged in freeze-etching and those doing cytochemistry of frozen sections. For the expert, this book is the equivalent of an informative tour of distant laboratories. Here they will learn of the latest "tricks" in freezing biological materials, direct viewing of frozen specimens, freeze-sectioning, freeze-drying, freeze-substitution, low-temperature embedding and freeze-fracturing.

The book will also be very helpful for the non-expert. It is a clear and accurate survey of a large number of methods, almost all of which are illustrated and wisely accompanied by detailed description and discussion. Although the amount of detail may appear excessive, success with low-temperature methods often hinges on apparently minute points. Freezing and freeze-etching methods are particularly well covered, and attention is given to tissue differences and important principles of sampling and interpretation that are often overlooked in "methods" books. Throughout, the authors provide critical comparisons and sufficient background and theory so that readers can decide for themselves what they can hope to gain by using low-temperature methods, and which method should be selected from the many available.

The book is abundantly illustrated but the graphics are often disappointing. The authors' attempt to be comprehensive is admirable, but the introductory diagrams are sometimes so cluttered with details of alternative pathways that many readers will have the impression of being given an unabridged chart of intermediary metabolic pathways before even having been introduced to the tricarboxylic acid cycle. An appendix listing suppliers of commercially available equipment is included, but

it is not as useful as it could have been: there are too few listings under specific categories and little attention has been given to suppliers outside Europe.

A breakthrough in low-temperature electron microscopy has been the demonstration that biological specimens can be vitrified. Although some of the techniques for the preparation and direct viewing of vitrified specimens at high resolution are included, the coverage here is neither as well organized nor as complete as this promising area deserves. To be fair, these methods are very recent; but it is regrettable that Robards and Sleytr did not have more of an opportunity to comment on these advances, for they have otherwise produced a discerning and up-to-date reference work. □

Daniel Branton is Higgins Professor in the Department of Cellular and Developmental Biology, Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

Graded grains

Peter D. Moore

Numerical Methods in Quaternary Pollen Analysis. By H.J.B. Birks and A.D. Gordon. Academic:1985. Pp.317. \$59, £45.

Pollen Records of Late-Quaternary North American Sediments. Edited by Vaughn M. Bryant Jr and Richard G. Holloway. American Association of Stratigraphic Palynologists Foundation, PO Box 819047, Dallas, Texas 75381, USA:1985. Pp.426. \$35.

ONE of the most valuable features of pollen grains as palaeoecological material is their sheer abundance. Combine this with their capacity to survive well in peats and lake sediments, and we have in fossil pollen assemblages a splendid opportunity to use numerical methods for their analysis. This was immediately apparent to von Post when he first came upon pollen in peat and it has been exploited by all subsequent research workers in the area. Until the 1960s numerical analysis was restricted to the expression of pollen counts in terms of percentage values, but since then there has been rapid growth in ways of treating data in order to improve their value to the palaeoecologist. Among those who have led this field of enquiry, few have contributed more to its development than the authors of *Numerical Methods in Quaternary Pollen Analysis*.

In their book, Birks and Gordon present a critical review of the current state of each aspect of Quaternary palynology to which numerical techniques have been applied, commencing with the most obvious, the presentation of pollen data. The

use of binomial and multinomial models, as originally described by Mosimann, is explained in some detail, though, as a basis for establishing homogeneity within a section of the stratigraphic profile, ordination methods are preferred. This involves the classification of samples, using spatial models, on the basis of dissimilarity indices.

This approach provides a number of opportunities to the palynologist, one of which is the establishment of objectively determined zones. These are subdivisions of the pollen profile into relatively homogeneous subunits for the sake of simplicity. In the early days a subjective approach to this process was considered perfectly adequate, since zones are simply aids to description of pollen proportions in time (that is, vertical space in the sediments). Recently many research workers have questioned the need for such zones at all, because they have led to a great deal of confusion in the past, while others feel that each of the main pollen taxa should be zoned independently. One wonders, therefore, whether sophisticated methods of zonation based on tests of homogeneity are necessary, but the practice is unlikely to do any harm and it may give confidence in the location of lines of maximum change.

Much more useful, in my opinion, is the application of the same group of techniques to follow the course of development of pollen assemblages through time, to compare one sequence with another and to set a fossil sequence in the context of modern pollen studies. Here the numerical approach comes into its own and many of those who doubted the value of these methods in the past, myself among them, have now become regular users.

Several variants of ordination methods are currently available for this type of work and they are well known to phytosociologists. These techniques are well explained by Birks and Gordon, and their application to palaeoecology is set out with clarity and enthusiasm. Many waverers have already been converted to their use and this book will be a means of initiating yet more palynologists into an increasingly fruitful type of analysis. Altogether, no palynological laboratory can afford to be without this compendium.

The second book under review, a manual of results rather than methods, is a long overdue synthesis of the vegetation history of North America through the late Pleistocene and Holocene. It is divided into geographical regions, ranging from Mexico to Alaska and Texas to Quebec, each dealt with by a different author. In each chapter representative pollen diagrams are reproduced and the variation in vegetation history is described, together, where appropriate, with assessments of climatic variations and human impact.

The emphasis is naturally upon vegeta-

tion change rather than methodology, but there are some helpful discussions of techniques such as the interpretation of alluvial pollen data in the drier regions, as in New Mexico, and the development of the isopoll vegetation maps in the north-east. The bibliographies supplied with each chapter are extensive and are a splendid point of entry to the relevant literature.

The timing of this volume is somewhat surprising, coming as it does on the heels of the two-volume *Late Quaternary Environments of the United States* (University of Minnesota Press/Longman, 1984). There is obviously a fair degree of overlap with that work, but the book reviewed here is confined to pollen data and is thus more detailed in vegetation descriptions and in its attention to regional variation.

Bringing so much information into one volume must have been a formidable undertaking and it has been brought off successfully only by careful planning and editing. Apart from finding a wide audience in North America, the book will be welcomed by palynologists on this side of the Atlantic who, perhaps, may be inspired to perform a similar service for Europe. □

Peter D. Moore is Reader in Ecology in the Department of Biology, King's College London (KQC), 68 Half Moon Lane, London SE24 9JF, UK.

"OUTSTANDING"

THE ATOMIC COMPLEX by BERTRAND GOLDSCHMIDT

An accurate, complete and fascinating worldwide political history and personal memoir of nuclear energy — from the development of the bomb in World War II to today's nuclear energy complex and proliferation problems.

"READS EASILY"

Goldschmidt, a leading French scientist turned international statesman, reviews half a century of political moves, countermoves, international intrigue and manipulation. *The Atomic Complex* is a pragmatic look at the nuclear world today, carefully examining the terror nuclear weapons represent, but at the same time stressing the benefits of nuclear energy.

Order this 500 page
best seller today.

\$31 Hardbound \$24 Softbound

**American Nuclear Society
555 North Kensington Avenue
La Grange Park, IL 60525 USA**

Done with a will

Richard Passingham

The Brain Machine: The Development of Neurophysiological Thought. By Marc Jeannerod. Translated by David Urion. Harvard University Press: 1985. Pp. 171. \$16.95, £14.50.

THE subtitle of *The Brain Machine* says that it deals with *The Development of Neurophysiological Thought*. It doesn't. In the original French the subtitle is *Physiologie de la Volonte*, and indeed the book turns out to be about the history of neurophysiological speculation about willed movement.

In Descartes' time people were credited with a will but animals were not. Descartes correctly conceived of the brain as a machine, although his conception of its workings was quaint. However, he also believed that by some mysterious process the machine was controlled by an indwelling soul.

Later physiologists, such as Hughlings Jackson, suggested a class division within the brain: based on the observation that damage to the motor cortex interfered with willed movements while leaving many automatic movements intact, the idea was that the lower centres catered for our animal needs, while the higher neocortical centres supervised our voluntary activities. Certainly, it is true that patients with such damage report that it is their subjective impression that moving the affected limb requires them to summon up an effort of will.

Jeannerod argues that "the elucidation of the term 'will', and the multiple concepts it covered, without doubt constitutes the central problem in comprehending the physiologic bases of the different aspects of movement" (p. 20), and in the various chapters of his book he covers some of these concepts. One issue is how a movement is triggered. We carry out some movements in response to external prompts, but there are others that we perform at our own direction and at our own pace. In accounting for spontaneous movement, the author mentions the spontaneous rhythms of the electroencephalogram and the oscillators proposed by von Holst. He also describes studies in which slow potentials are recorded from the cortical motor areas while subjects make simple movements. These potentials differ in their pattern and timing when it is the subject rather than the experimenter who decides when the movements are to be made.

A second idea is that of intention. Jeannerod argues that if a movement is intended the brain knows "what will happen as a result of its own action" (p. 97); that is, the brain is aware of its own commands

and can therefore compare what happens with what was intended. A copy of the command to the motor system is sent to other centres, but if a movement occurs which is not intended there will be no such copy.

There is, too, a third notion: "Human action, when it is voluntary, intentional, or propositional, has the distinction of being dominated by language, in the sense that language precedes the action and is an integral part of the representation" (p. 121). More generally there must be a plan; people can try out possible actions in their heads before performing them. Recent technology has allowed us to detect evidence of such planning. When brain regions increase their activity they increase their metabolic requirements, and we can use the flow of blood to an area as a measure of its metabolic activity. It is now possible to take these measurements from the outside of the skull and thus to investigate the activity of brain regions in human subjects while they carry out mental work. If subjects are asked to rehearse mentally a set of finger movements without actually performing them, activity can be detected in the supplementary motor area.

Jeannerod's book gives the history of these and other ideas. He includes chapters on such topics as the natural history of the soul; hierarchy and integration of movements; the motor cortex; the physiology of spontaneity; representation, plan, programme; and action, the force of self-organization. He ranges widely, his theme taking him from animal spirits to electrical conduction, and from spinal reflexes in frogs to apraxia in the hospital clinic, and also into the realms of European philosophers such as Schopenhauer, Cassirer and Merleau-Ponty. Because the thrust of the book is primarily historical, Jeannerod says little of modern experiments in modern physiology. But the history of ideas on this subject is intriguing, and makes an altogether fascinating story.

The book is illustrated with line drawings, some of which are delightful: for example, the author reprints Descartes' splendid drawing of the mechanism for muscular contraction, a naive diagram by an alchemist of mental function and a beautiful drawing of a brain dissection done in 1827. The text has been well translated by David Urion, and it is to his particular credit that if the book were handed to someone who did not know they might fail to guess the language of the original. □

Richard Passingham is a Lecturer in the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.

• Recently published by Raven is *The Citadel of the Senses*, by Macdonald Critchley, a collection of 30 short essays on episodes and personalities in the "prescientific age of neurology". The book costs \$32.50, and will be reviewed in a future issue of *Nature*.

The mean mass density of the Universe

P. J. E. Peebles

Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544, USA

Observations of galaxy clustering indicate that the mean mass density is only about one-third of that predicted by the popular Einstein-de Sitter cosmological model. Theory and observation can be reconciled if there has been large-scale segregation of galaxies from mass, but the gravitational instability of the expanding universe would tend to destroy such a segregation. Thus either conventional ideas about galaxy formation are wrong, or the Einstein-de Sitter model is not a useful approximation.

THE mass distribution of the universe is approximately uniform on a large scale, and can be described by two parameters: the mean mass density ρ and the expansion rate H . Conventional general relativity gravity theory indicates that if ρ is large enough the expansion eventually stops and the universe collapses to a singularity, while for small ρ the universe continues to expand forever. This conventional theory has two defects: it fails to explain why the large-scale mass distribution should be so close to uniform and it requires that the expansion can be extrapolated back to a singularity at a definite time in the past. As will be indicated below, Guth's¹ inflation model offers a partial solution to both problems. In the usual interpretation, inflation requires that the mean mass density ρ be equal to the critical value that separates universes that eventually collapse from those that expand forever, so that ρ has a definite value fixed by the expansion rate, H . However, as will be described, estimates of ρ based on derived masses of clusters of galaxies and counts of these objects are lower than the critical value by a factor of ~ 3 . It is impressive that the observed density is so close to the expected critical value, but it is also apparent that the discrepancy is real. We can avoid the problem by recalling that the gravitational acceleration of a hollow spherical mass is zero inside the sphere. Thus estimates of the mass of a cluster of galaxies based on the accelerations of the galaxies in the cluster would miss any mass on the outskirts beyond the observed galaxies. The 'biased galaxy formation' model, in which galaxies form in islands surrounded by a sea of dark mass, can thus reconcile the observed mass density with the larger theoretically expected value. However, there is yet another consideration. The universe is gravitationally unstable, which means that gravity is drawing galaxies together to form an ever larger hierarchy of clusters. Because different forms of mass fall at the same gravitational acceleration, the gravitational instability draws mass together with galaxies. Thus if galaxies formed in tight islands, then the collections of islands formed later by gravity would be expected to yield a higher apparent mean mass density. The fact that this is not observed suggests that either galaxies formed in a different way from the standard picture, in islands large enough that all available observations miss the sea, or else that ρ really is smaller than the critical value.

The idea that the mean mass density might be expected to equal the critical value predates the inflation model. The historical development of this idea is reviewed in the next section, in order to show why the value of ρ has received so much attention and has so strongly influenced the direction of research in cosmology. There follows a review of observational estimates of ρ , which shows that the measurements are now sufficiently detailed to merit serious attention. Possible ways to reconcile the observations with the higher critical mass density are dis-

cussed next. Finally, I indicate how we might most simply adapt our ideas on cosmology if a low value of ρ is accepted.

Historical remarks

Cosmological models. In the general relativity cosmological model the rate of expansion, measured by the Hubble constant, H , and the mean mass density, ρ , are related by the equation

$$H^2 = \frac{8}{3}\pi G\rho \pm \frac{c^2}{(aR)^2} + \frac{\Lambda}{3} \quad (1)$$

The radius of curvature of space (measured at a fixed cosmic time, t) is $a(t)R$ (where R is the cosmic scale factor), Λ is Einstein's cosmological constant, and the dimensionless expansion parameter $a(t)$ is proportional to the mean proper distance between conserved particles (such as galaxies). G is the universal gravitational constant, and c is the speed of light. The mass density term is often written as $\frac{8}{3}\pi G\rho = H^2 \rho/\rho_c = \Omega H^2$, where $\rho_c = 1.88 \times 10^{-29} h^2 \text{ g cm}^{-3}$ and $H = 100 h \text{ km sec}^{-1} \text{ Mpc}^{-1}$. The length unit is 1 Mpc (megaparsec) = $3.1 \times 10^{24} \text{ cm} \approx 3 \times 10^6 \text{ light yr}$. The dimensionless parameter h reflects the uncertainty in the Hubble constant; it is thought to be in the range $0.5 \leq h \leq 1$. The critical density is ρ_c . If $\Lambda = 0$, as is usually assumed, then if $\rho < \rho_c$ the universe continues expanding indefinitely, and if $\rho > \rho_c$ the universe eventually collapses. The density parameter is the ratio $\Omega = \rho/\rho_c$ of the mean mass density to the critical value.

Arguments for $\rho = \rho_c$. The value of Ω is of course an observational question, but one gets a useful guide to its significance by reviewing the theoretical arguments which have been advanced in support of $\Omega = 1$.

The cosmological model with $\Lambda = 0$ and $\Omega = 1$ is named after Einstein and de Sitter^{2,3}, who argued as follows: in the original static Einstein cosmological model the presence of a non-zero mean mass density ρ (with a reasonable pressure) requires that Λ and space curvature both assume specific non-zero values. By 1932 it was recognized that the static model is a special case; a positive mean mass density generally requires evolution in time (expansion or contraction of the universe). Einstein and de Sitter noted that in evolving models with non-zero mass density there is no requirement that Λ or space curvature be non-zero, and suggested that these terms might be dropped. De Sitter recognized that if $\Lambda = 0$ and $R^{-2} = 0$ it would require a rather large mean mass density, but he argued that this is not inconsistent with the observations. However, his estimates of the number density of galaxies and the mean mass per galaxy exceed the values found by Hubble⁴, and in the more recent work summarized below.

In a second argument, one observes that the mass density term in equation (1) $\frac{8}{3}\pi G\rho$, varies at least as rapidly as $(a(t))^{-3}$,

the curvature term varies as $(aR)^{-2} \propto (a(t))^{-2}$ and the last term is $\Lambda/3 = \text{constant}$. The three terms thus vary with time in three different ways. The values of these terms were presumably set as initial conditions in the very early universe when classical general relativity theory became a useful approximation and when the expansion parameter $a(t)$ was very much smaller than it is now. If $\Omega \neq 1$, at least two of the terms on the right-hand side of equation (1), density and one other, make an appreciable contribution to the present expansion rate, H . But because these terms vary with $a(t)$ in different ways, in the very early universe the two terms that are comparable now would have been vastly different, but by the highly special factor needed to make the terms comparable just now. This is an unlikely arrangement. The advantage of the Einstein-de Sitter model is that the one term we know is present, the mass density, always dominates. This attractive argument based on the undesirability of coincidences is not easy to find in the literature before the introduction of the inflation model, although Dicke^{5,6} has advanced it for 25 years.

Wheeler⁷ has long argued that Mach's principle, which states that inertial motion is defined by the mean motion of mass in the universe, sets boundary conditions for general relativity, requiring that space be closed, like the surface of a sphere. This eliminates Minkowski space (which is the limit to which an open expanding universe with $\Lambda = 0$ tends as $t \rightarrow \infty$), in which test particles have inertia but there is no background matter to define inertial motion. If $\Lambda = 0$ a spatially closed universe requires $\Omega > 1$, which means the density is high.

In Guth's¹ inflation model, the expansion of the universe can be traced back to a phase transition before which the energy density of space would have had a large, nearly constant term (which behaves like a large Λ -term) that would have caused space to expand as an exponential function of time (instead of the power-law expansion of the usual models). If this exponential 'inflationary' phase lasted long enough it could have stretched all characteristic lengths, in whatever chaos the inflation started from, to values larger than the maximum distance we can see, $cH^{-1} \approx 3,000 h^{-1}$ Mpc. That would account for the fact that we see such very small density gradients around the sky at depths $\approx cH^{-1}$. But then it would have to follow that the radius of curvature $|aR|$, is also much larger than cH^{-1} , so the curvature term would have to be negligibly small in equation (1). The coincidence argument mentioned above says that Λ ought to be negligible. Particle physicists observe that Λ has the units of mass to the fourth power (with suitable factors of G , c and Planck's constant). A reasonable mass from particle physics yields a value of Λ suitable for the inflationary phase but more than 100 orders of magnitude larger than is now observationally acceptable¹. The only other 'reasonable' value is $\Lambda = 0$, which brings us to $\Omega = 1$.

Finally, galaxy formation seems to be easier to understand if $\Omega \approx 1$ because if $\Omega < 1$ the kinetic energy of expansion dominates the gravitational potential energy, thus suppressing the growth of mass concentrations by gravity⁸, so that to get the observed degree of small-scale clustering of mass in galaxies and clusters of galaxies we would have to postulate larger initial mass-density fluctuations in the early universe, which would cause fluctuations in the microwave background that are uncomfortably large compared with the observational limits⁹.

Exotic matter. Evidence for a low mean mass density in baryons comes from the computed production of elements in the hot Big Bang. It is difficult to see how the abundance of deuterium in our stellar neighbourhood could have been produced in stars. If an appreciable part of the deuterium were to have been produced in the early universe in the conventional hot Big Bang cosmology it would require a mass density of baryonic matter that (with a value of H in the standard range) would imply $\Omega \leq 0.1$ (ref. 10). However, as noted by Chapline¹¹ and Gott *et al.*¹⁰, that constraint would be lifted if the mass of the universe were dominated by non-baryonic matter such as small black

holes or massive neutrinos; the list of such possible exotic forms of dark matter is now quite long¹². Thus it appears that if $\Omega = 1$ the mass of the universe must be dominated by exotic matter other than the material out of which we and stars and the bright parts of galaxies are made. There are some prospects for the experimental detection of exotic matter if it is present in our neighbourhood of the Galaxy¹³.

Segregation of galaxies from mass. It was well established by the early 1970s that the 'seen' mass in the dense central parts of galaxies is well below what is required in the Einstein-de Sitter model, but there were already indications that there is considerably more mass in 'massive faint extended haloes' of galaxies¹⁴. Gott *et al.*¹⁰, arguing partly from the nucleosynthesis constraint on the baryon density, concluded that $\Omega \approx 0.06 \pm 0.02$. Ostriker *et al.*¹⁵ and Einasto *et al.*¹⁶ argued that there is in fact a lot of mass outside the bright parts of galaxies, the observations suggesting that the mass of a galaxy increases almost linearly with radius to at least ~ 1 Mpc, which would bring the density parameter to $\Omega \approx 0.2$. If the trend of mass with radius extrapolated to the mean separation between galaxies, $\sim 4 h^{-1}$ Mpc, it would bring Ω quite close to unity¹⁷; however, there has recently been considerable improvement in the dynamical estimates of masses on scales larger than that of a single galaxy and, as described below, it appears that the trend does not continue. If the mean mass per galaxy derived from the dynamics of groups and clusters were assigned to all galaxies it would yield $\Omega \approx 0.3$, quite close to the estimates of a decade ago.

Although dynamical estimates of Ω use whole-galaxy masses (because they consider systems of galaxies) they could nonetheless underestimate Ω if there were a smoothly distributed mass component that has little effect on clustering dynamics. In the now standard version of this idea it is argued that galaxy formation would be expected to be biased toward regions of higher than average mass density, where galaxies could more readily form and survive. The current interest in this idea largely developed out of a series of papers that appeared in published or preprint form in 1984. Kaiser's¹⁸ argument that clusters of galaxies may cluster more strongly than do galaxies themselves, because clusters are unusually dense spots that tend to have dense surroundings, was soon generalized to the suggestion that galaxies cluster more strongly than mass because galaxies form at unusually dense spots^{19,20}. Rees²¹ emphasized that it is easy to think of ways in which the formation of a galaxy could influence the formation of neighbours; and Davis *et al.*²² found that N -body model simulations of the motion of pressureless matter in an Einstein-de Sitter model would give reasonable results if galaxies were located at the high (2.5 standard deviations) mass density peaks. Biased galaxy formation is the framework for most recent papers on galaxy formation²²⁻²⁸.

Estimates of Ω

Three methods have yielded useful measures of Ω from galaxy clustering dynamics. We are in a small collection of galaxies, the Local Group, on the edge of a large cloud, the Local Supercluster (LSC), which is roughly centred on the Virgo cluster of galaxies. The mass concentration in the LSC produces an inward peculiar velocity field, v_v . We can hope to measure v_v and so derive the local peculiar gravitational field and hence the mass excess, δM , in the LSC. If this is a fair sample then the mean mass per galaxy in the LSC multiplied by the global mean number density of galaxies yields the global mean mass density^{29,30}. Recent estimates^{31,32} of v_v are in the range $150 \leq v_v \leq 450 \text{ km s}^{-1}$. The concentration of bright galaxies within our radius in the LSC is $\delta N/N = 2.2 \pm 0.3$, where N is the expected mean number of galaxies for a homogeneous distribution and δN the observed number minus N . The recession velocity of the Local Group from the Virgo cluster is $cz = HR_v - v_v = 980 \pm 50 \text{ km s}^{-1}$, where R_v is our distance from the centre of the mass concentration. (z is the redshift corresponding to this recession velocity.) If the material in the LSC is a fair sample,

so that the mass concentration is $\delta M/M = \delta N/N$, then these numbers yield³¹⁻³³

$$0.05 < \Omega_{\text{LSC}} < 0.45 \quad (2)$$

The large spread results mainly from the uncertainty in v_v . To obtain $\Omega = 1$, v_v would have to be $\approx 900 \text{ km sec}^{-1}$, but it is unlikely that the observations could have missed such a large value. With the numbers given above, $\Omega = 1$ requires that the mass constraint in the LSC be $\delta M/M = 0.5-1$, which is a factor of 2-4 less than $\delta N/N$, and it is difficult to believe that the counts could be that far off. The connection between v_v and $\delta M/M$ assumes either linear perturbation theory or spherical symmetry. In an interesting cautionary example, Bushouse *et al.*³⁴ find that in N -body models the value of Ω derived from the mean v_v may be wrong by 30-50% (after correction for nonlinearity; A. Melott, personal communication). However, what such N -body models mainly show is a model velocity field considerably rougher than observed, indicating that the real mass distribution somehow stays a good deal smoother than in the models. One can estimate the possible effect of departure from spherical symmetry another way, by asking what transverse velocity, v_t , of our Galaxy (and its neighbours) relative to the LSC would be needed if the present gravitational acceleration of the LSC were balanced by the present centrifugal acceleration due to v_t . If $\Omega = 1$ the result is

$$v_t^2 = \frac{1}{2} \frac{\delta M}{M} (HR)^2 \quad (3)$$

With $\delta M/M = \delta N/N$ and the above numbers this yields $v_t \approx 1,200 \text{ km sec}^{-1}$, which is about three times bigger than could be allowed, so it is difficult to see how the estimate could be spoiled by departures from radial motion. Thus the evidence is that if $\Omega = 1$, the mass concentration $\delta M/M$ in the LSC is less than about half the galaxy concentration $\delta N/N$.

For the discussion of biased galaxy formation it will be useful to have an estimate of what the LSC would have been like when galaxies formed. The numbers here assume the Einstein-de Sitter model, with $\Omega = 1$, and assume pressure may be neglected in estimating the growth of clustering. As quasars are seen at redshifts z (where $1+z$ is the ratio of observed to emitted wavelength) ≈ 4 , it is reasonable to assume that the sites of galaxy formation were fixed at a redshift z_i no less than about 5. The mass concentration in the LSC can be extrapolated back from its present value, $\delta M/M \approx 0.5-1$, to $(\delta M/M)_i \approx 0.10$ at $z_i = 5$, and $(\delta M/M)_i$ scales as $(1+z_i)^{-1}$. The galaxy concentration extrapolates back to $(\delta N/N)_i \approx 1$, very nearly independent of z_i . The mean number density of bright galaxies is $\langle n \rangle \approx 0.02 h^3 \text{ Mpc}^{-3}$ (ref. 35). (The bright galaxies are those that contribute 75% of the net mean luminosity per unit volume.) The number of bright galaxies N_{is} within our radius in the LSC is the product of $\langle n \rangle$ with the volume and $(1+\delta N/N)$: $N_{is} \approx 600$. Thus in the biased galaxy formation model the Local Supercluster would have originated as a moderately prominent island of ~ 600 bright galaxies, with $(\delta N/N)_i \approx 1$, but with a modest mean-mass-density enhancement, $(\delta M/M)_i \approx 0.1$.

The second method of estimating Ω is based on the mean mass distribution around pairs of galaxies. If galaxy redshifts are used to derive distances assuming pure Hubble flow (that is, distance $= cz/H$), then random relative motions make the galaxy clustering pattern appear elongated along the line of sight. The size of the effect is a measure of the relative velocity of pairs of galaxies^{36,37}. This measure has been applied to three moderately deep samples of galaxies selected by apparent magnitude and to two samples of more or less isolated pairs of galaxies³⁷⁻⁴⁰. The root-mean-square (r.m.s.) relative line-of-sight velocity dispersion, σ , of pairs of galaxies as a function of the projected separation, r_p , is found to increase slowly with increasing separation at $5 \text{ kpc} \leq r_p \leq 1 \text{ Mpc}$; typical values are $\sigma = 200 \pm 20 \text{ km sec}^{-1}$, $hr_p = 10 \text{ kpc}$; $\sigma = 300 \pm 75 \text{ km sec}^{-1}$, $hr_p = 1 \text{ Mpc}$. The first number agrees with what is expected from

observed velocities within galaxies at $hr \approx 10 \text{ kpc}$ (where r is the radial distance from the galaxy centre). The flatness of $\sigma(r_p)$ is an extension of the nearly flat rotation or velocity dispersion curves observed in spiral and some elliptical galaxies^{41,42}.

The relative motions of close galaxy pairs must be balanced on average by gravity, because otherwise the clustering pattern observed at $hr \leq 1 \text{ Mpc}$ would dissolve within a Hubble time (H^{-1}). If we model the mass distribution as the linear sum of spherical haloes with mass (within r) $\delta M(<r)$ centred on each galaxy, then the condition for statistical support of the clustering pattern at small r , where typically two galaxies are involved, is $\delta M(<r) \approx \sigma^2 r / G$. (This assumes isotropic velocities. If orbits tended to be radial then, as discussed below, this would reduce δM and Ω .) We can compare δM with the known mean number of galaxies in excess of random within a distance r of a galaxy:

$$\delta N(<r) = \langle n \rangle \int \xi_{gg}(r) d^3r = r^{3-\gamma} \quad (4)$$

where $\xi_{gg} = (r_0/r)^\gamma$, $\gamma = 1.77$ and $hr_0 = 5.4 \pm 0.3 \text{ Mpc}$. $\xi_{gg}(r)$ is a two-point correlation function measured from galaxy counts, and $\langle n \rangle$ is the global mean galaxy number density. The power-law model is a good approximation at $hr \leq 10 \text{ Mpc}$ ^{38,43,44}. We see that because σ varies roughly as $r^{0.1}$, δM and δN scale with r in about the same way, as $r^{1.2}$. This invites us to consider the assumption that whatever gathered galaxies together to make the present clustering pattern gathered mass in the same proportion, so that $\xi_{gp} = \xi_{gg}$ and therefore $\delta M(<r)/\delta N(<r) = \langle \rho \rangle / \langle n \rangle$, where $\langle \rho \rangle$ is the global mean mass density. The result of solving the above equations for $\Omega = \langle \rho \rangle / \rho_c$ is

$$0.3 \leq \Omega_{gg} \leq 0.6 \quad (5a)$$

This applies at $hr \leq 100 \text{ kpc}$, where the galaxy pair typically has no bright close neighbours.

At $hr \approx 1 \text{ Mpc}$, where a galaxy pair typically has many ($N_{is} \approx 10$) bright neighbours that can contribute to the mean relative gravitational acceleration, we assume that each galaxy has mass m and then use the known statistics of the galaxy distribution to compute the mean relative acceleration. This yields^{38,39}

$$0.1 \leq \Omega_{gg} \leq 0.3 \quad (5b)$$

The third dynamical estimate is based on the mean distribution of galaxies and mass around the rich compact clusters of galaxies catalogued by Abell⁴⁵. Let $n(r)$ and $\rho(r)$ be the mean values of the galaxy number density and the mass density at distance r from an Abell cluster and averaged across a fair sample of clusters, and write $n(r) = \langle n \rangle [1 + \xi_{cg}(r)]$ and $\rho(r) = \langle \rho \rangle [1 + \xi_{cp}(r)]$, where $\langle n \rangle$ and $\langle \rho \rangle$ are the large-scale mean values, and ξ_{cg} and ξ_{cp} are analogous to ξ_{gg} and ξ_{gp} , but for clusters. The Licck⁴⁶ galaxy counts around the positions of Abell richness class 1 clusters yield⁴⁷

$$\xi_{cg} = \left(\frac{7.0 \text{ Mpc}}{hr} \right)^{2.5} + \left(\frac{12 \text{ Mpc}}{hr} \right)^{1.7} \quad (6)$$

at $0.5 \leq hr \leq 40 \text{ Mpc}$. Gravity will have caused $n(r)$ and $\rho(r)$ to have grown considerably since galaxies formed. If gravity pulled in mass and galaxies alike, then at large enough r , ξ_{cp} must approach ξ_{cg} , in which case the known ξ_{cg} fixes $\rho(r)/\langle \rho \rangle$. If we can estimate $\rho(r)$ from dynamics we have an estimate of $\langle \rho \rangle$ ⁴⁷.

The following assumes $\xi_{cg} = \xi_{cp}$ at $r \approx 2 h^{-1} \text{ Mpc}$. This avoids the dense central parts of the cluster (where the galaxies are gas-poor so the mean mass per galaxy might be anomalous) but is still close enough to the centre to permit a reasonable estimate of $\rho(r)$. The Struble-Rood catalogue (M. F. Struble & H. J. Rood, unpublished) lists 17 Abell clusters of richness class 1 in which enough redshifts are known to permit estimates of the line-of-sight galaxy velocity dispersion σ_0 ; the r.m.s. value is $\langle \sigma_0^2 \rangle^{1/2} = 830 \pm 80 \text{ km sec}^{-1}$. The measurements of σ_0 are concentrated well within the Abell radius ($hr_A = 1.5 \text{ Mpc}$). For the

four clusters in which the variation of velocity dispersion with radius has been fairly well measured (the Coma, Perseus, Hydra I and Virgo clusters), the $\sigma(r)$ estimates decrease with increasing r outside a possible slight central minimum⁴⁸⁻⁵¹. These results suggest that $\sigma(r)$ at $hr \approx 2$ Mpc is down from the central value by a factor of 1.3 to 1.5.

If the galaxy orbits in the cluster are isotropic, with velocity dispersion $\sigma(r)$ varying only slowly with radius, the mass density needed for statistical equilibrium is $\rho(r) = \sigma(r)^2 / (2\pi Gr^2)$. This implies $\rho \approx r^{-2}$, consistent with the above estimate of $\xi_{cg}(r)$. The above numbers at $hr = 2$ Mpc give

$$0.2 < \Omega_{cg} < 0.35 \quad (7)$$

This assumes an isotropic velocity distribution. A spherical system is described by two components of dispersion, σ_r in the radial direction and σ_t in one transverse direction. Following White⁵², we assume that at $hr \approx 2$ Mpc, σ_r and σ_t have reached constant values and the galaxy number density scales as a power law, $n \propto r^{-\epsilon}$. Then the equilibrium gravitational acceleration, g is

$$g = \frac{GM(<r)}{r^2} = \frac{2\sigma_t^2 + (\epsilon - 2)\sigma_r^2}{r} \quad (8)$$

and the line-of-sight velocity dispersion is $\sigma^2 = \sigma_r^2/\epsilon + \sigma_t^2(1 - 1/\epsilon)$. It is a reasonable approximation to take $\epsilon = 2$; this simplifies the equations to

$$\frac{GM}{r\sigma^2} = \frac{4\sigma_t^2}{\sigma_t^2 + \sigma_r^2} \quad (9)$$

Equation (7) assumes σ_t . If orbits were purely circular, with $\sigma_t \gg \sigma_r$, this would double M and Ω_{cg} , but it is hard to see how this could come about. If clusters formed by collapse it would tend to produce radial orbits, with $\sigma_r > \sigma_t$, which would depress Ω_{cg} from the value in equation (7).

If Ω were unity, clusters would have to be stronger concentrations of galaxies than of mass, and we would expect that gravity would attract to the cluster material rich in mass per galaxy, making the mass per bright galaxy $m(r)$ increase with increasing radius r in the cluster. Attempts to measure $m(r)$ using galaxy velocity dispersion data alone are complicated by the uncertainty in how the velocity anisotropy parameter σ_t^2/σ_r^2 varies with r ; for the Coma cluster Kent and Gunn⁴⁸ could only conclude that they saw no need to assume that m varies with r . Cowie *et al.* (L. L. Cowie, J. Henriksen and R. Mushoksky, in preparation) have shown that the constraints on $m(r)$ are improved if one also imposes the condition that the gravitational field must balance the pressure-gradient force in the plasma in the cluster. Using the integrated cluster X-ray spectrum as a constraint on the distribution of plasma temperatures in the cluster, the rather accurately measured plasma density as a function of r , and the galaxy velocity dispersion data, they conclude that $m(r)$ has to be a decreasing function of r in the Coma cluster. The constraints are weaker for the Perseus cluster, but there is no evidence that $m(r)$ increases with increasing r , contrary to what we would have expected to see if Ω were unity.

We avoid this problem if we assume that cluster galaxies were created in large islands with roughly uniform mass per galaxy within the island. (The gradient in $m(r)$ in the Coma cluster is then given the usual interpretation, that galaxies near the centre suffered earlier loss of gas, which would have suppressed formation of younger stars and so lowered the galaxy luminosities.) The minimum number of bright galaxies per island is the mean number within our matching radius of $2 h^{-1}$ Mpc: $N_{is} \geq 70$.

It may be possible to extend this constraint by using the fact that the concentration of galaxies around rich clusters extends well beyond $2 h^{-1}$ Mpc. For example, the mean number of

galaxies within distance $r = 20 h^{-1}$ Mpc of an Abell cluster is about twice the number expected for a homogeneous distribution⁴⁷. If there were a like concentration of mass within $r = 20 h^{-1}$ Mpc the gravitational attraction of the mass would cause the galaxies at $r \approx 20 h^{-1}$ Mpc to have an appreciable streaming motion toward the cluster (relative to the Hubble flow), as in the LSC. When galaxy distances are estimated from redshifts using Hubble's law, $r = cz/H$, this streaming motion would make the galaxy distribution around a cluster tend to appear flattened along the line of sight⁵³. Application of this test would require a large redshift sampling program, which now seems feasible and would be of considerable interest.

A more indirect constraint on the mass distribution might be mentioned. The new Center for Astrophysics (CfA) survey catalogues 1,100 galaxy redshifts in a field 6° wide by $\sim 100^\circ$ long⁵⁴. The redshift map shows holes with diameters $\approx 30 h^{-1}$ Mpc, which is about what would be expected, given the clustering seen on smaller scales. The surprising result is that the holes are so empty and that the galaxies tend to distribute themselves so smoothly along the walls of the holes. If the mass were distributed like the galaxies and if Ω were close to unity this would be a strongly unstable arrangement, gravity causing mass to drain along the walls toward intersections, as is seen in the numerical N -body model simulations of the pancake model for galaxy formation⁵⁵. Thus if $\Omega = 1$ it is doubtful that the mass is in the walls; it is more reasonable to assume that there is a general segregation of galaxies from mass on a scale of $\sim 30 h^{-1}$ Mpc. If so then we have no problem in reconciling the low apparent densities in equations (2), (5) and (7) with a high mean mass density, but we substantially increase the demands on a theory for biased galaxy formation. If $\Omega < 1$, then gravity is suppressed relative to the general expansion, and a mass distribution like the galaxy distribution in the CfA redshift map might reasonably be expected to persist.

The direct way to measure Ω is through the classical cosmological model tests, as discussed many years ago by Tolman⁵⁶ and others. This approach has not received much attention in recent years because it has become apparent that galaxy evolution can mask the effects one is looking for, but the classical tests may be due for a renaissance. Loh and Spillar^{57,58} have shown that a measurement of the joint distribution, $f(m, z)$, of galaxy apparent magnitude and redshift is feasible and substantially reduces the evolution problem. The standard tests depend on the curvature of a function of one variable, such as apparent magnitude as a function of redshift, and a given curvature can be produced by many reasonable combinations of galaxy evolution and cosmological effects. The function $f(m, z)$ depends on two variables, which offers a much better chance of separating the effects of galaxy evolution and cosmology. The current result²⁸ is $0.6 < \Omega < 2.0$ if $\Lambda = 0$. This is encouraging for the Einstein-de Sitter model but it should be considered a preliminary result because it assumes all galaxy luminosities evolve at the same fractional rate: the possible effect of evolution of the shape of the galaxy luminosity function has yet to be tested.

Models for biased galaxy formation

Three points are worth noting from the dynamical estimates of Ω described above. First, the independent estimates in equations (2), (5) and (7) yield consistent results, $\Omega \approx 0.3$. If Ω were unity and the estimates were biased because mass clusters less strongly than galaxies, then we would have expected the Ω estimates to increase with increasing clustering scale (as measured by N_{is}). Second, if Ω were unity then galaxies would have to be placed on islands of tens to hundreds of bright galaxies each; a bias extending to nearest neighbours alone would not do. Third, whatever put the galaxies in islands cannot have perturbed the mass very much. If the Local Supercluster formed at redshift $z_i \approx 5$, the initial mass concentration would have been $\delta M/M \approx 0.1$, which is 10% of the original galaxy concentration. Even if

galaxies were put in place by non-gravitational forces it would require a special arrangement to keep the resulting mass perturbation so small.

One model for biased galaxy formation starts from the idea¹⁹⁻²¹ that the first generation of stars that formed in the densest spots in the primeval mass distribution might have suppressed star formation elsewhere (perhaps blast waves from supernovae push the gas out of protogalaxies that have not yet managed to collapse to stars), so that bright galaxies would have survived only at rare extreme positive mass density fluctuations. Because the peaks of the mass distribution can be more strongly clustered than is the general mass distribution¹⁸ (assuming the distribution is a random gaussian process with positive autocorrelation function), galaxies would be segregated from mass in the desired way. In a common implementation of this idea one assumes that a galaxy forms wherever the primeval mass density, smoothed through a window with size typical of a galaxy, exceeds the mean by more than a fixed multiple, $\nu \approx 3$, of the standard deviation. This prescription has been studied in some detail under the assumptions that the mass of the universe is dominated by exotic dark matter with negligible primeval pressure and that the primeval mass distribution is a random gaussian process (random phases) with power spectrum P (square of the Fourier transform) that scales as $P \propto \langle |\rho_k|^2 \rangle \propto k$, where k is the wavenumber. (The assumption $P \propto k$ comes from the expected spectrum of quantum fluctuations resulting from inflation¹.)

Analytical¹⁹ and N -body model²² calculations show that these assumptions with the above prescription for the sites of galaxy formation make galaxies cluster more strongly than mass, as desired. However, one can use the same analytical computation to show that the bias does not extend much beyond the nearest neighbour galaxy. Thus this model would help account for the low value of Ω derived from relative motions of close pairs of galaxies (equation (5)), but would not help with the estimates based on larger clustering scales (if $\Omega = 1$).

To extend the scale of biasing one can assume that the spectrum, P , of density fluctuations increases with increasing length scale, rather than decreasing as in the above model. In this case we have to discuss in more detail whether the prescription for galaxy formation sites is reasonable. Three points should be considered.

First, the prescription uses the mass distribution averaged over a typical scale of a galaxy. Unless one assumes that the spectrum P is cut off on scales below that of a galaxy, one would expect that the first generation of stars formed in clusters much smaller than a galaxy and well before objects the size of a galaxy could have formed (because mass fluctuations tend to increase with decreasing length scales). It is hard to see how such pre-galaxy star clusters could be eliminated once formed. (The star clusters once formed could coalesce later to form galaxies.) This suggests that our criterion for star-forming regions should be the peaks in the primeval mass distribution averaged over scales well below that of a galaxy. But the analytical argument mentioned above indicates that this would suppress biasing on the scales of galaxies and larger.

Second, as mentioned above, the biased galaxy formation model requires that the Local Supercluster form at a mass excess $(\delta M/M)_i < 0.1$ and with a galaxy excess $(\delta N/N)_i \approx 1$. I doubt that it is reasonable to expect that a 10% change in the ambient mass density could have caused a factor of two change in the survival of galaxies, because if protogalaxies were easily destroyed, survival surely would depend on many factors other than ambient mass density. Examples are the degree of sub-clustering in a protogalaxy, which determines how many star clusters form early and so resist dispersal, and the degree of sheltering of a protogalaxy by its neighbours. These additional random variables would tend to suppress the sensitivity of the survival probability to one variable, the ambient mass density.

Third, one might have thought that the dispersal effect which

is supposed to inhibit galaxy formation would decrease with increasing distance from a new galaxy. This would tend to suppress galaxy clustering, leading to a reverse bias in which galaxies are less strongly clustered than mass, which would make the true value of Ω less than the dynamical estimates. To get the desired sign of bias, we would have to assume that tens to hundreds of bright galaxies form in islands and become robust enough to withstand disruption by their immediate neighbours, while protogalaxies well outside the islands must so lag in star formation that they can be destroyed by blast waves from the islands. My impression is that no very convincing way to do this has been proposed.

Another approach to biasing starts not from the statistics of the primeval mass distribution, but from the expected rearrangement of the mass distribution as stars form and die in supernova explosions. In this model, discussed by Ostriker and Cowie⁵⁹ and by Ikeuchi⁶⁰, a blast wave from supernovae in an early star cluster compresses the surrounding gas into a ridge, which can fragment into new star clusters whose blast waves can make still larger ridges. Islands of galaxies would be produced wherever the last and largest ridges form. As noted above, if we adopt $\Omega = 1$ and we accept the constraint on the baryon density from nucleosynthesis, we must assume that the mass of the universe is dominated by exotic, weakly interacting dark matter. The dark matter would be left behind when the blast waves gather the hydrogen to make galaxies, but then would participate in clustering once gravity takes over, giving the desired partial segregation of galaxies from mass.

This model requires two special arrangements. First, the explosions must be adjusted so as to fill space without making most of the matter so hot it cannot cool and contract to stars. Second, we must find some other way to account for the tendency of galaxies to appear in clusters and of clusters to appear in superclusters, because there is no way known to make the blast waves travel distances comparable to the sizes of superclusters. If these objects formed by the gravitational growth of primeval mass density fluctuations, as in the first model discussed above, it would require a special coincidence to account for the observations that galaxies and clusters of galaxies fit quite well into a scale-invariant clustering hierarchy^{8,43}. It remains to be seen how serious these two problems are.

There are still other ways in which galaxies might have been segregated from mass. If the baryons were initially distributed in a wispy fashion in a sea of exotic dark matter, galaxies could form only in the wisps. A roughly similar effect might be obtained if galaxy formation were triggered by a network of massive cosmic strings. Both models are *ad hoc* for our purpose, but not without supporting arguments^{61,62}. Another possibility is that the mass of the universe is dominated by gravitational radiation, or by weakly interacting particles with velocities that have always been high enough to resist the gravitational attraction of the mass concentrations in galaxies. However, either would significantly increase the mean mass density at high redshift, which would speed expansion and so spoil the nucleosynthesis calculations⁶³. An ingenious variant of this scheme proposes that galaxies and clusters of galaxies formed by gravitational clustering of massive, weakly interacting, unstable particles⁶⁴. These particles are then supposed to have decayed into low-mass, high velocity, weakly interacting particles, leaving behind islands of galaxies in a nearly smooth sea of dark matter. This is not without cosmological problems^{65,66}. For example, if clusters of galaxies lost two-thirds of their mass, which would make the apparent density parameter derived from clustering dynamics ≈ 0.3 , the clusters would expand by a factor of ~ 3 in radius. That means the clustering pattern would have been stronger in the past, which is contrary to the usual assumption that the clustering pattern has been growing to the present epoch. This would substantially increase the density fluctuations needed to make clusters of galaxies form by gravitational instability, and thus would increase the problem of reconciling the gravitational

instability picture with the isotropy of the microwave background^{9,66}.

Cosmology with low mass density

As the above models as presently understood are not very persuasive, we might consider the alternative: cosmological models with low mass density. The Einstein-de Sitter model could fail because we are assuming the wrong physics. An example is Milgrom's conjecture, that we are using the wrong gravity theory to compute masses⁶⁷. A less spectacular mode of failure is the possibility that general relativity theory is valid, and that space sections have zero curvature, in accordance with inflation, but that the cosmological constant is non-zero⁶⁸. This is distasteful, because of the coincidence argument mentioned at the beginning of this review, but such arguments must be examined closely. For example, suppose $\Omega = 0.2$, $h = 0.75$, $R^{-2} = 0$, and assume that initial conditions for classical cosmology were set at the end of inflation at redshift $z_{c1} = 10^{27}$. Expansion starts from $a = 1$ at $z = z_{c1}$. Four important epochs are: equal relativistic and non-relativistic mass densities, $a_1 = 10^{23.6}$; decoupling of baryons from radiation, $a_2 = 10^{24.0}$; equal values of the Λ and ρ terms in equation (1), $a_3 = 10^{26.8}$; and development of astronomy on Earth, $a_4 = 10^{27.0}$. It is curious that a_1 is relatively close to a_4 , and striking that a_1 is so close to a_2 ; but surely it is no more startling that a_3 might be so close to a_4 . It might be counted as a distinct advantage that if $\Omega \approx 0.1$, the mass indicated by clustering dynamics is at least roughly consistent with the baryon density allowed by current estimates of element production in the hot Big Bang⁶³ without invoking exotic matter.

Conclusions

The Einstein-de Sitter cosmological model has been considered to be the most reasonable choice by the majority of those who cared to adopt an opinion in advance of a convincing observational test. We are now entering a very interesting period as the tests improve and, among the dynamical methods, fail to support the Einstein-de Sitter model. The observations can be reconciled with our expectations by assuming that galaxies formed in islands in a more nearly uniform sea of dark mass, but my impression is that we have seen no model that does this without a considerable degree of special arrangement. This impasse is a part of the normal process of development of an active but still immature subject, and will probably be resolved soon. It seems likely that in the next few years, as redshift surveys proceed, we will be able to improve considerably the measures of the large-scale mass distribution and so sharpen the constraints on Ω and the related problem of how galaxies form. It is somewhat more optimistic to hope for a tight constraint on Ω from the classical cosmological tests⁵⁸. More optimistic still, but not unreasonably so, is the hope that observations of objects at high redshift will help show us how galaxies formed⁶⁹. Meanwhile, the tension between what is observed and what we think ought to be is a most useful stimulus to research, as long as we remember that it is only a stimulus.

I thank P. Campbell, Len Cowie, Bob Dicke, George Efstathiou, John Huchra, Steve Kent, Ed Loh, Ardrian Melott, Jerry Ostriker, Jeff Peterson, Martin Rees, Dave Schramm, Joe Silk, Chris Thomson and Dave Wilkinson for stimulating discussions. This research was supported in part by NSF.

- Guth, A. *Ann. N. Y. Acad. Sci.* **422**, 1-14 (1984).
- Einstein, A. & de Sitter, W. *Proc. natn. Acad. Sci. U.S.A.* **18**, 213-214 (1932).
- Einstein, A. *Sur la Structure Cosmologique de l'Espace*, 108 (Hermann, Paris, 1933).
- Hubble, E. *Realm of the Nebulae*, 188-189 (Yale University Press, 1936).
- Dicke, R. H. *Gravitation and the Universe*, 62 (American Philosophical Society, Philadelphia, 1970).
- Dicke, R. H. & Peebles, P. J. E. in *General Relativity: an Einstein Centenary Survey* (eds Hawking, S. W. & Israel, W.) 504-517 (Cambridge University Press, New York, 1979).
- Misner, C. W., Thorne, K. S. & Wheeler, J. A. *Gravitation*, 549 (Freeman, San Francisco, 1973).
- Peebles, P. J. E. *Astrophys. J.* **189**, L51-L53 (1974).
- Bond, J. R. & Efstathiou, G. *Astrophys. J.* **285**, L45-L48 (1984).
- Gott, J. R., Gunn, J. E., Schramm, D. N. & Tinsley, B. M. *Astrophys. J.* **194**, 543-554 (1974).
- Chapline, G. F. *Nature* **253**, 251-252 (1975).
- Pagels, H. R. *Ann. N. Y. Acad. Sci.* **442**, 15-32 (1984).
- Goodman, M. W. and Witten, E. *Phys. Rev. D* **31**, 3059-3063 (1985).
- Peebles, P. J. E. *Physical Cosmology*, 116 (Princeton University Press, 1971).
- Ostriker, J. P., Peebles, P. J. E. & Yahil, A. *Astrophys. J.* **193**, L1-L4 (1974).
- Einasto, J., Kaasik, A. & Saar, E. *Nature* **250**, 309-310 (1974).
- Peebles, P. J. E. *Ann. N. Y. Acad. Sci.* **375**, 159 (1981).
- Kaiser, N. *Astrophys. J.* **284**, L9-L12 (1984).
- Bardeen, J. in *Inner Space/Outer Space* (eds Kolb, E. W. et al.) (University of Chicago Press, in the press).
- Kaiser, N. in *Inner Space/Outer Space* (eds Kolb, E. W. et al.) (University of Chicago Press, in the press).
- Rees, M. J. in *Clusters and Groups of Galaxies* (eds Mardirosian, F. et al.) 485-497 (Reidel, Dordrecht, 1984).
- Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. *Astrophys. J.* **292**, 371-394 (1985).
- Faber, S. M. in *Dark Matter in the Universe* (eds Kormendy, J. & Knapp, G.) (Reidel, Dordrecht, in the press).
- Rees, M. J. *Mon. Not. R. astr. Soc.* **213**, 75P-81P (1985).
- Schaeffer, R. & Silk, J. *Astrophys. J.* **292**, 319-329 (1985).
- Dekel, A. & Silk, J. *Astrophys. J.* **303**, 39-55 (1986).
- Melott, A. L. & Fry, J. N. *Astrophys. J.* **305** (in the press).
- Kaiser, N. & Davis, M. *Astrophys. J.* **297**, 365-370 (1985).
- Silk, J. *Astrophys. J.* **193**, 525-527 (1974).
- Peebles, P. J. E. *Astrophys. J.* **205**, 318-328 (1976).
- Davis, M. & Peebles, P. J. E. *A. Rev. Astr. Astrophys.* **21**, 109-130 (1983).
- Tammann, G. A. & Sandage, A. *Astrophys. J.* **294**, 81-95 (1985).
- Hoffman, G. L. & Salpeter, E. E. *Astrophys. J.* **263**, 485-500 (1982).
- Bushouse, H., Melott, A. L., Centralia, J. & Gallagher, J. S. *Mon. Not. R. astr. Soc.* **217**, 7P-11P (1985).
- Kirschner, R. P., Oemler, A., Schechter, P. & Shectman, S. A. *Astr. J.* **88**, 1285-1300 (1983).
- Geller, M. J. & Peebles, P. J. E. *Astrophys. J.* **184**, 329-341 (1973).
- Peebles, P. J. E. *Astr. J.* **84**, 730-734 (1979).
- Davis, M. & Peebles, P. J. E. *Astrophys. J.* **267**, 465-482 (1983).
- Bean, A. J. et al. *Mon. Not. R. astr. Soc.* **205**, 605-624 (1983).
- Peebles, P. J. E. *Science* **224**, 1385-1391 (1984).
- Rubin, V. C. *Science* **220**, 1339-1344 (1983).
- Fabricant, D. & Gorenstein, P. *Astrophys. J.* **267**, 535-546 (1983).
- Groth, E. J. & Peebles, P. J. E. *Astrophys. J.* **217**, 385-405 (1977).
- Stevenson, P. R. F. et al. *Mon. Not. R. astr. Soc.* **213**, 953-969 (1982).
- Abell, G. O. *Astrophys. J. Suppl. Ser.* **3**, 211-288 (1958).
- Shane, C. D. & Wirtanen, C. A. *Univ. Calif. Publ. Lick Obs.* **22**, 1-60 (1967).
- Seldner, M. & Peebles, P. J. E. *Astrophys. J.* **214**, L1-L4 (1977).
- Kent, S. M. & Gunn, J. E. *Astr. J.* **87**, 945-971 (1982).
- Kent, S. M. & Sargent, W. L. W. *Astr. J.* **88**, 697-708 (1983).
- Ftaclas, C., Fanelli, M. N. & Struble, M. F. *Astrophys. J.* **282**, 19-28 (1984).
- Fuchs, B. & Marterne, J. *Astr. Astrophys.* **113**, 85-93 (1982).
- White, S. D. M. *Mon. Not. R. astr. Soc.* **195**, 1037-1056 (1981).
- Peebles, P. J. E. *Large-Scale Structure of the Universe*, 297-299 (Princeton University Press, 1980).
- De Lapparent, V., Geller, M. J. & Huchra, J. P. *Astrophys. J.* **302**, L1-L5 (1986).
- Melott, A. *Mon. Not. R. astr. Soc.* **202**, 595-604 (1983).
- Tolman, R. C. *Relativity, Thermodynamics and Cosmology*, (Clarendon Press, Oxford, 1934).
- Loh, E. D. & Spillar, E. J. *Bull. Am. astr. Soc.* **61**, 488-489 (1984).
- Loh, E. D. & Spillar, E. J. *Astrophys. J.* (in the press).
- Ostriker, J. P. & Cowie, L. L. *Astrophys. J.* **243**, L127-L131 (1981).
- Ikeuchi, S. *Publ. astr. Soc. Japan* **33**, 211-222 (1981).
- Stecker, F. W. *Nucl. Phys. B* **252**, 25-36 (1985).
- Turock, N. & Schramm, D. N. *Nature* **312**, 598-599 (1984).
- Yang, J., Turner, M. S., Steigman, G., Schramm, D. N. & Olive, K. A. *Astrophys. J.* **281**, 493-511 (1984).
- Turner, M. S., Steigman, G. & Krauss, L. M. *Phys. Rev. Lett.* **52**, 2090-2093 (1984).
- Efstathiou, G. *Mon. Not. R. astr. Soc.* **213**, 29P-34P (1985).
- Silk, J. & Vittorio, N. *Phys. Rev. Lett.* **54**, 2269-2272 (1985).
- Milgrom, M. *Astrophys. J.* **270**, 365-389 (1983).
- Peebles, P. J. E. *Astrophys. J.* **284**, 439-444 (1984).
- Djorgovski, S., Spinrad, H., McCarthy, P. & Strauss, M. A. *Astrophys. J.* **299**, L1-L5 (1985).

Black smokers, massive sulphides and vent biota at the Mid-Atlantic Ridge

P. A. Rona*, G. Klinkhammer†, T. A. Nelsen*, J. H. Trefry‡ & H. Elderfield§

* NOAA/AOML, 4301 Rickenbacker Causeway, Miami Florida 33149, USA

† Department of Earth, Atmospheric and Planetary Sciences, MIT, Cambridge, Massachusetts 02139, USA

‡ Department of Oceanography and Ocean Engineering, Florida Institute of Technology, Melbourne, Florida 32901, USA

§ Department of Earth Sciences, University of Cambridge, Cambridge CB3 0EZ, UK

The discovery of black smokers, massive sulphides and vent biota in the rift valley of the Mid-Atlantic Ridge demonstrates that this assemblage of hydrothermal phenomena is not limited to intermediate-to-fast-spreading oceanic ridges. Hydrothermal exchange processes may thus be important at the ridges which extend through the Atlantic Ocean and western Indian Ocean, comprising more than half of the 55,000-km global length of seafloor spreading centres.

AN assemblage of hydrothermal phenomena including active black smokers, inactive chimneys, massive sulphides and vent biota has been discovered at a site¹ in the rift valley of the Mid-Atlantic Ridge near latitude 26°08' N (Fig. 1). This is the first report of such an assemblage at a slow-spreading oceanic ridge (half-rate ≤ 2 cm yr⁻¹). The association of black-smoker-type hydrothermal venting, chimney-like edifices precipitated by the venting solutions, and vent biota was previously known only at sites on intermediate-to-fast-spreading ridges in the eastern Pacific Ocean, including the East Pacific Rise²⁻⁴, the Guaymas Basin of the Gulf of California⁵ and the Juan de Fuca Ridge^{6,7}. Vent biota have been reported from sites at cold seeps in the Gulf of Mexico^{8,9} and at the axis of the Mid-Atlantic Ridge ~310 km south of the present study site¹⁰, and vent clams were tentatively identified near the present site¹.

The black smokers and associated phenomena were found using a combination of methods to sense far- and near-field properties of the water column and sea floor on a July–August 1985 cruise of the NOAA ship *Researcher* as part of the NOAA Vents program. Shipboard measurements of elevated values of dissolved manganese¹¹ and light scattering due to suspended particulate matter¹² in the water column delineated an area of hydrothermal activity. A systematic survey of that area with a deep-towed instrumented sled located active vents by measuring near-bottom water temperature anomalies and recording images of the sea floor¹³. We report here preliminary results of the seafloor imaging, near-bottom water temperature measurements and analyses of dredged hydrothermal precipitates¹⁴. These data complement other data sets pertaining to water chemistry¹¹, suspended particulate matter^{12,15}, and bottom sediment¹⁶.

Instrumentation

Twelve transects were made across the study site with an instrumented sled. The sled was towed at an average speed of 1.8 km h⁻¹ within 10 m of the sea floor, using an acoustic pinger to monitor and a winch to adjust altitude above the sea floor. The position of the sled was determined every 15 s to an accuracy of ± 20 m relative to a net of four long-baseline (5–10 km) bottom-mounted acoustic transponders. The geographical positions of the acoustic transponders were determined using shipboard satellite navigation (Global Positioning System for three of the four transponders); the acoustic navigation net was converted to geographical coordinates with an estimated accuracy of ± 150 m. In addition to the acoustic pinger and a relay transponder, the sled carried a 35-mm still camera system programmed to fire at 7.5-s intervals (inoperative 50% of the time due to malfunction); a continuously operating video camera system with recorder; a 1.5-m-long vertical array of four temperature sensors (thermistors accurate to ± 0.01 °C) which recorded sequentially for 2 s every 10 s; and a pressure sensor with

recorder from which pressures were converted to sled depth¹⁷ (accuracy ± 2 m) and summed with pinger altitude for seafloor depth.

Geological setting

The discovery was made within the previously known TAG (Trans-Atlantic Geotraverse) hydrothermal field¹⁸⁻²¹, which is located on a 10-km segment of the east wall of the rift valley between small-offset (<10 km) right-lateral transform faults²². The two walls of the rift valley of a slow-spreading ridge are normally nearly straight and subparallel to the bathymetric axis of the rift valley. At the TAG hydrothermal field the east wall projects westward as a broad salient which narrows the rift valley floor from a full width of ~6 km to only 3 km, coincident with a bathymetric high along the rift valley axis (Fig. 1). Thus, the east wall may be close to the subadjacent intrusive heat sources that provide the increased thermal gradients which drive the upwelling limb of a sub-seafloor hydrothermal convection system. The sea floor is spreading asymmetrically at half-rates (averaged over 10 Myr) of 1.1 cm yr⁻¹ to the west and 1.3 cm yr⁻¹ to the east^{23,24}, indicative of asymmetric crustal accretion to the east^{1,21}.

The east wall rises from the rift valley floor near 4,000 m depth to the top of the rift mountains at ~2,000 m in a series of steps formed by underlying fault blocks. The steps are up to hundreds of metres in horizontal and vertical dimension, and tilt eastward away from the rift valley at angles up to 30° (refs 25, 26). Previous work identified a zone of hydrothermal phenomena between 2,400 and 3,100 m on the east wall¹. These phenomena comprise layered deposits and surface stains of manganese oxide, iron oxide, iron hydroxide and iron silicate²⁷, as well as anomalous values of temperature¹ and helium concentration²⁸ in the near-bottom water, and metal enrichment in the sediment²⁹. Analysis of metals in cores recovered from thin (<1.5 m) sediment between 3,100 m and the rift valley floor revealed layers with anomalously high contents (calculated on a carbonate-free basis) of copper (>1,000 p.p.m. (parts per 10⁶ by weight)), iron (>8%) and zinc (>1,000 p.p.m.) at the surface and 30 cm beneath the surface, attributed to past and recent episodes of high-temperature black-smoker-type venting^{1,29}. The layered hydrothermal precipitates, surface stains and near-bottom water temperature anomalies exhibit a linear distribution along fault zones, primarily between fault blocks that trend subparallel to the rift valley. The fault zones are inferred to tap hydrothermal sources, channel upwelling hydrothermal solutions and focus their discharge^{20,21}.

Black smokers, sulphides and vent biota

The black smokers, massive sulphides and vent biota occur at the juncture between the floor and east wall of the rift valley at

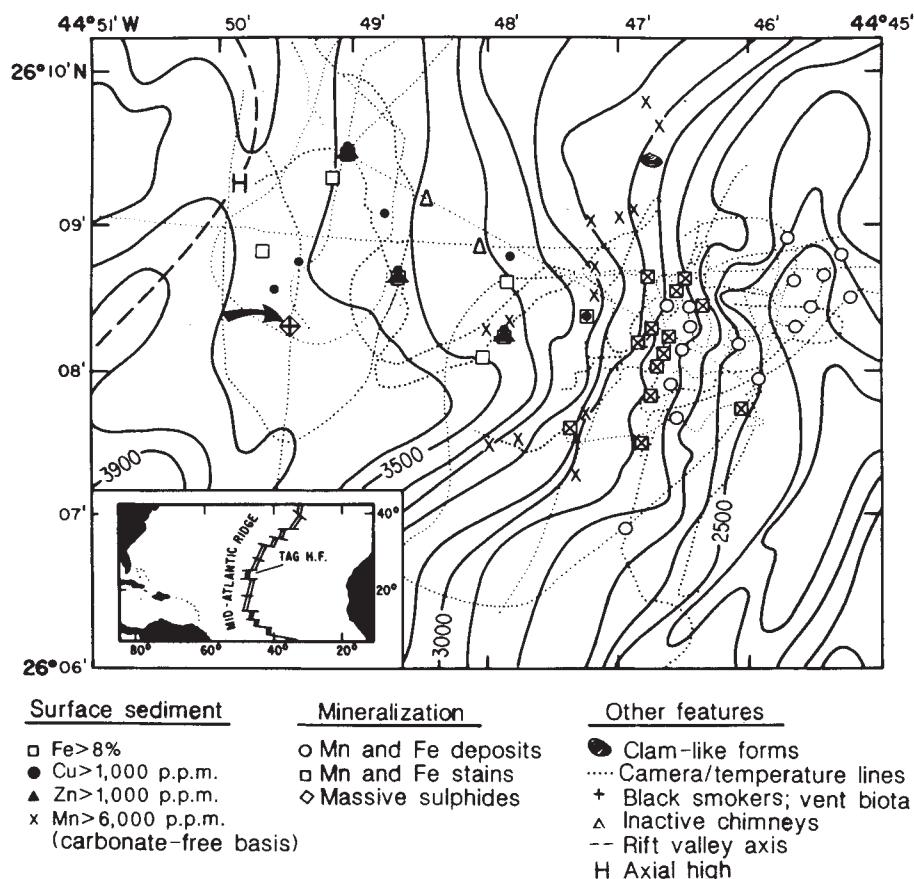
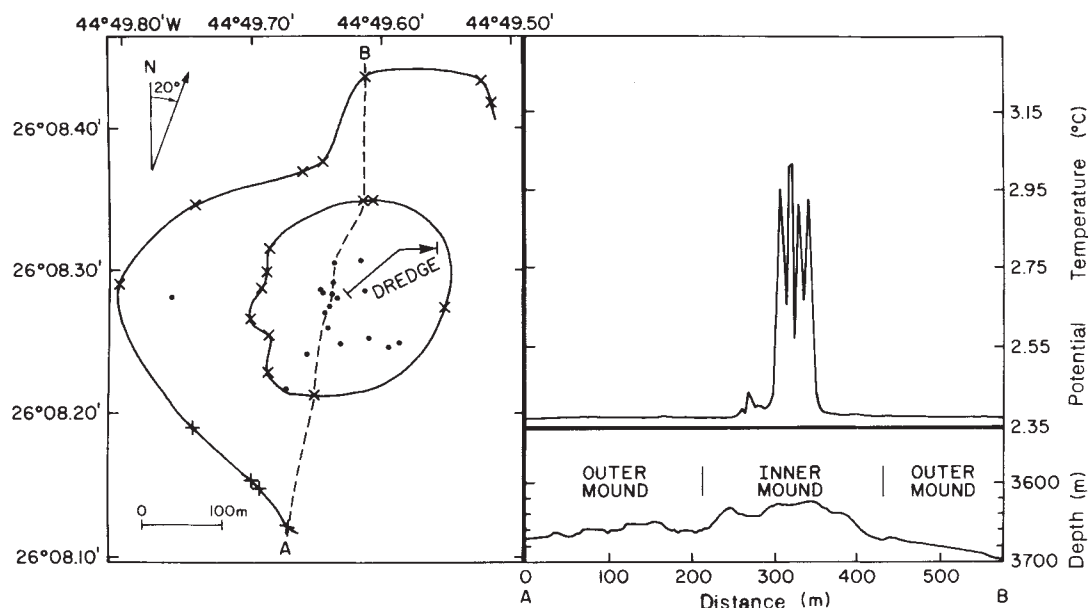


Fig. 1 Bathymetric map (100-m isobaths) of the TAG hydrothermal field, showing mineralization and other features on the east wall of the rift valley based on the present and previous work. The mound where black smokers, vent biota and massive sulphides occur is marked with an arrow. Camera-temperature lines shallower than 3,500 m were made in 1982; those deeper than 3,500 m were made in 1985; several of the 1985 lines crossing the black smokers were too closely spaced to plot. Inset shows location in the central North Atlantic Ocean.

Fig. 2 Characteristics of the mound where black smokers, vent biota and massive sulphides occur at the juncture between the east wall and floor of the rift valley (Fig. 1), as delineated by camera-temperature-depth transects. Left, plan view showing boundaries of concentric inner and outer mounds (crossings marked 'x'), distribution of vents (dots), azimuth of axis of rift valley (020°), and positions of transect A-B and a dredge. Lower right, bathymetric profile across the mound along transect A-B. Upper right, potential temperature profile in the water column within 10 m of the sea floor along transect A-B; the four central spikes correspond to black smokers.



water depths of 3,620–3,700 m (Fig. 1). They lie between the previously delineated hydrothermal zone 3.7 km up-slope to the east and the bathymetric axis of the rift valley 1.5 km down-slope to the west. The axial high of the rift valley lies 2 km up-slope to the north-west. The camera-temperature-depth tows delineated a compound mound with concentric inner and outer portions, each with an elliptical shape in plan view (Fig. 2). The inner mound lies between water depths of ~3,620 and 3,675 m and is up to 55 m higher than the outer mound, which lies between 3,675 and 3,700 m. The inner mound is ~250 m wide along azimuth 020° by 220 m wide along azimuth 110°; the outer mound is ~580 m wide along azimuth 020°.

A representative transect shows principal features of the compound mound (Fig. 2, transect A-B). The sea floor surrounding the mound is covered by basalt talus interspersed with sediment. The talus ranges from cobble- to boulder-sized, including angular fragments and whole pillows; some of the pillows may be in place. The sediment is thin (<1.5 m), light tan to pale red, mottled with light-coloured spots several centimetres in diameter made by burrowing organisms (~5 spots m⁻²), fine-grained and resuspends in the wake of the sled. At the perimeter of the outer mound the talus is predominantly boulder-sized and sharply angular. Some of these talus blocks display regular, repetitive layers ~5 cm thick, which resemble successive basalt flows

Table 1 Petrography and mineralogy of representative samples

Sample no.	Petrography*	Pyrite (FeS ₂)		Marcasite (FeS ₂)		Chalcopyrite (CuFeS ₂)		Mineralogy sphalerite (ZnS)		Anhydrite (CaSO ₄)		Gypsum (CaSO ₄ ·2H ₂ O)		Silica (SiO ₂)	
		Content†	Wt %‡	Content	Wt %	Content	Wt %	Content	Wt %	Content	Wt %	Content	Wt %	Content	Wt %
1	Spongy-textured inter-grown yellow pyrite and black sphalerite	Major	82.0§	Minor	—	Very low	0.8	Minor	13.6	ND	—	ND	—	Very low	0.4
2	Fragment of chimney wall 2 cm thick with inner zone of chalcopyrite and pyrite and outer zone of sphalerite	Minor	24.2§	Low	—	Major	63.1	Very low	4.7	ND	—	ND	—	Very low	0.4
3	Coarsely crystalline anhydrite (density 3.07) with fine pyrite crystals and an orange mineral (molysite?) on surfaces and in cavities	Very low	0.7	ND	—	Very low	2.9	Very low	0.04	Major	68.0	Low	—	Very low	0.9
4	Fragment of chimney wall similar to sample 2	Very low	—	ND	—	Major	86.1	Very low	1.4	ND	—	ND	—	Very low	0.4
5	Fine-grained, unconsolidated, cohesive pyrite particles with some euhedral crystals of pyrite	Major	92.8§	Low	—	ND	—	ND	—	ND	—	ND	—	Very low	0.4

Major, >25 wt %; minor, 10–25 wt %; low, 5–10 wt %; very low, <5 wt %; ND, not detected.

* Polarizing and scanning electron microscopy.

† X-ray diffraction.

‡ Wt % estimated from mineralogy and major element analysis by B. W. Haynes, US Bureau of Mines.

§ Wt % FeS₂ includes both pyrite and marcasite.

|| Wt % CaSO₄ includes both anhydrite and gypsum.

ponded in lava lakes, such as are observed at the axis of the East Pacific Rise³⁰. Alternatively, the layering may have been produced by bedded volcanoclastics. The sea floor within the outer mound is covered by angular talus fragments, basalt flows and interspersed sediment. The talus consists of a mixture of angular basalt fragments and reddish-yellow layered rocks inferred to be hydrothermal precipitates. Basalt pillows are present, both detached in the talus and attached as part of lava flows partially buried by talus and sediment. The sediment is red, mottled with abundant light-coloured spots at burrows (~10 spots m⁻²), fine-grained and resuspends in the wake of the sled. Typically one large fish (≤25 cm long) was observed within each 100 m of horizontal advance of the sled.

The inner mound contrasts with the outer mound as regards substrate, hydrothermal activity and biota. The substrate is predominantly composed of reddish-yellow layered rocks which form irregular protuberances: ledges, dome-shaped masses and columns with horizontal and vertical dimensions of the order of metres. During tows across the inner mound the sled occasionally hit bathymetric highs while travelling at an altitude of ~5 m above the sea floor; thus the towing procedure systematically missed imaging the tops of the higher protuberances which may have been chimneys.

Shimmering water emanates from broad areas of fractured rocks inside the boundary of the inner mound. By analogy with temperature measurements made by submersibles in shimmering water at the Galapagos spreading centre³¹ and the Guaymas Basin³², this water could be up to 200 °C hotter than the ambient water potential temperature of 2.37 °C (Fig. 2). Towards the centre of the inner mound, suspended particulate matter in the water column increases in quantity, and changes from discrete, neutrally buoyant white flocks inferred to be pieces of bacterial mats, to wisps of white smoke, to wisps followed by plumes of black smoke.

The black smokers emanate from fractures in the layered rocks (Fig. 3), rather than from distinct chimneys. The buoyant plumes of black smoke appear to rise more slowly than those investigated on the East Pacific Rise^{33,34}, which ascend at rates between 1 and 5 m s⁻¹. At the TAG site, the rate of plume ascent was estimated to be <1 m s⁻¹ from videotapes and from the

down-current inclination of plumes (Fig. 3) in currents flowing south at several centimetres per second (estimated from particle movement at the time of observation). Particle loading in the plumes is high, as evidenced by the fact that the sea floor behind the plumes is obscured by the density of the black smoke (Fig. 3). Also, the camera sled was coated with fine, cohesive black particulate matter after tows through the smokers. Effects of the hydrothermal plumes were detected up to 400 m above the sea floor, as anomalies in dissolved manganese¹¹ and suspended particulate matter^{12,15}.

At least 11 black smokers clustered near the centre of the inner mound were seen during various transects (Fig. 2). The highest potential temperature anomaly measured from the sled at an altitude of 5 m above the sea floor was 3.86 °C. This temperature is 1.43 °C above ambient, and is associated with a



Fig. 3 A black smoker photographed from a video image recorded at an altitude of 5 m over the central portion of the inner mound (Fig. 2). The hydrothermal solutions are being discharged from seafloor fractures and the plume is inclined down-current.

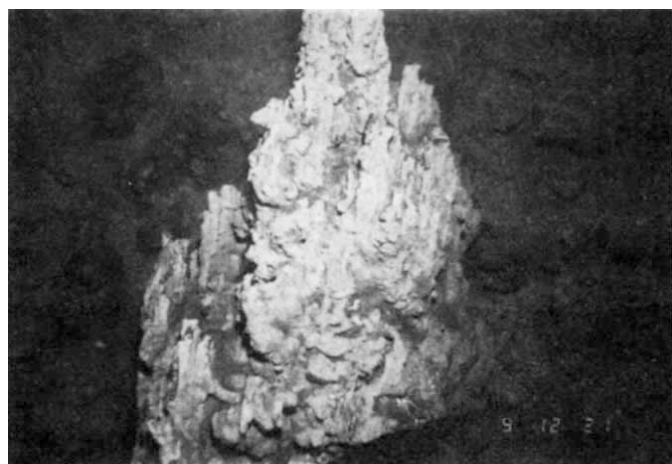


Fig. 4 A toppled chimney ~4 m from base to top, lying on the sea floor in an area of inactive chimneys at 3,500 m water depth on the lower portion of the east wall of the rift valley (Fig. 1). The sea floor around the chimney is littered with fragments of layered rocks with the appearance of hydrothermal precipitates and thin sediments with patchy black mineralized stains.

gradient of $0.21\text{ }^{\circ}\text{C m}^{-1}$ (warming downward) over the 1.5-m vertical length of the temperature sensor array. The highest water temperature previously reported from a sled towed over hydrothermal vents in the Pacific Ocean was $0.18\text{ }^{\circ}\text{C}$ (ref. 35), because the anomalies are attenuated by rapid heat dissipation in moving water. Direct measurements made from a submersible in similar Pacific vents have recorded water temperatures up to $400\text{ }^{\circ}\text{C}$ (ref. 36).

The colour of the sediment interspersed between seafloor protuberances changes from red near the margins of the inner mound to black next to the black smokers. The black sediment and adjacent rocks exhibit a spectacular glitter on the videotape, interpreted as reflections from faces of euhedral pyrite crystals that settle from the black smoke. Euhedral pyrite crystals were identified in the black sediment that coated the instrument sled after tows through the black smokers (Tables 1, 2; sample 5).

The biota exhibits systematic changes from the margins to the centre of the inner mound. Translucent anemones several centimetres in diameter appear at the boundary between the outer and inner mound. Such anemones are the predominant sessile benthic invertebrate on the inner mound, with the densest population in shimmering water. Sparse translucent worms tentatively identified as vestimentiferans (M. L. Jones, personal communication) lie recumbent on rocks in limited areas closer to the centre: the thickets of vestimentiferan worms conspicuous at Pacific hydrothermal sites at shallower depths (2,600 m) were not observed. Occasional white patches (~1 m in diameter) on the rocks appear to be bacterial mats. Thousands of shrimp bask in beds around the black smokers and are highly mobile. About 200 of these shrimp, ranging in length from 3.8 to 5.1 cm, were recovered in a dredge with rocks from the inner mound (Fig. 2); the shrimp were preserved and are being studied⁴¹. A few crabs identified as brachyurans (A. B. Williams, personal communication) were present on rocks near the shrimp beds. Large fish (up to 25 cm long), including an eel-like form, were more abundant on the inner than on the outer mound, especially near the shrimp beds. Clams and other shelled organisms were not observed, although vesicomyid clams had been tentatively identified previously, in photographs from 3,000 m in the hydrothermal zone¹ (Fig. 1).

Area of inactive chimneys

A group of chimneys was observed on bottom images from two transects of an area at 3,500–3,600 m depth, located 2 km north-east of the mound (Fig. 1). One of the transects imaged toppled chimneys (Fig. 4) and related fragmental material at 3,500 m, with interspersed sediment marked by dark stains similar to mineralized stains observed higher on the east wall, there attributed to active hydrothermal seepage (Fig. 1). The other transect imaged 35 standing chimneys up to 5 m high along a

horizontal distance of 100 m at 3,600 m depth. The absence of near-bottom water temperature anomalies over either the toppled or standing chimneys indicates that they are inactive; however, the fresh appearance of the chimneys (Fig. 4) suggests that hydrothermal venting stopped quite recently. By analogy with chimneys of similar appearance at sites in the Pacific, it is likely that these chimneys contain massive sulphides.

Hydrothermal precipitates

A single dredge of the inner mound (Fig. 2) recovered 59 kg of hydrothermal precipitates without any basalt. The hydrothermal precipitates recovered are predominantly sulphides (66 wt % of dredged material); sulphates, oxides, hydroxides, silicates and carbonates were also recovered. Preliminary petrographic descriptions and mineralogical and compositional analyses of five representative samples were performed at the Avondale Research Center of the US Bureau of Mines (B. W. Haynes, personal communication). The consolidated sulphides are broadly classified as two types. The first type (93 wt % of sulphides recovered) is a spongy-textured intergrowth of yellow pyrite and black sphalerite with irregular layering and cavities

Table 2 Composition of representative samples (wt%, dry basis)

	Method*	1	2	Sample no. 3	4	5
Major elements						
Fe	1	38.4	30.5	1.20	22.3	43.2
Cu	1	0.28	21.9	1.00	29.8	0.008
Zn	1	9.15	3.13	0.03	0.95	0.02
Ca	1	0.04	0.04	31.2	0.04	0.04
Mo	1	0.03	0.02	<0.01	0.05	<0.01
Ni	1	<0.004	0.06	<0.004	0.04	<0.004
Pb	1	0.06	0.19	<0.01	0.25	0.02
Si	1	0.2	0.2	0.4	0.2	0.2
S	4	46.9	36.1	17.4	30.7	50.9
Minor elements						
Al	1	0.05	0.19	0.02	0.47	0.04
Ag	2	<0.03	<0.03	<0.03	<0.03	<0.03
Ba	1	0.03	0.01	<0.01	<0.01	0.03
Co	1	<0.001	0.003	<0.001	0.003	<0.001
K	3	0.08	0.11	0.20	0.09	0.19
Mg	1	<0.01	0.01	<0.01	<0.01	<0.01
Mn	1	<0.005	<0.005	<0.005	<0.005	<0.005
Na	2	0.25	0.33	0.05	0.15	1.21
Sr	1	<0.01	<0.01	0.21	<0.01	<0.01
Ti	1	0.06	0.09	0.05	0.12	0.01
V	1	<0.01	<0.01	<0.01	0.02	<0.01
H ₂ O	(105 °C)	0.33	0.32	0.23	0.42	1.16

* Method of analysis: 1, inductively coupled plasma spectrometer; J. Zink and C. Hammett, US Bureau of Mines; 2, semi-quantitative optical emission spectrographic analysis; 3, atomic absorption spectrometer; 4, sulphur analyser. Oxygen content undetermined.

(Tables 1, 2; sample 1). It is high in iron (38 wt %) and zinc (9 wt %) and low in copper (≤ 1 wt %). Oxidized crusts typically several millimetres thick are present on exposed surfaces. The characteristics of this predominant sulphide type suggest precipitation from moderate-temperature solutions discharging into open spaces and forming irregularly layered deposits of diverse shapes on the sea floor. The second type of sulphide (7 wt % of sulphides recovered) occurs in dense, curved plates up to several centimetres thick exhibiting distinct mineral zonation (Tables 1, 2; samples 2, 4). A relatively thick zone on the concave side of the plates is composed of chalcopyrite and pyrite; a thinner contiguous zone on the convex side is composed of sphalerite (Tables 1, 2; samples 2, 4). The convex surface has a yellowish-to reddish-brown oxidation crust several millimetres thick; the concave surface of some of the samples is coated with soft, black, botryoidal and reddish-brown, fine-grained iron and manganese oxides and hydroxides up to 1 cm thick. The copper content of these sulphides is high (22–30 wt %), and the zinc content is low (1–3 wt %). The characteristics of this second type of sulphide suggest precipitation in a chimney that vented high-temperature solutions. Unconsolidated sulphide particles that accumulated on the instrument sled during tows through the black smokers consist of cohesive particulate material primarily composed of pyrite (Tables 1, 2; sample 5), as described above.

Sulphates (~1 kg) were recovered in the form of coarsely crystalline aggregates of anhydrite with fine crystals of pyrite on surfaces and in cavities (Tables 1, 2; sample 3). The presence of anhydrite indicates seafloor temperatures of ~140 °C: at lower temperatures it would re-dissolve in sea water, at the ambient pressure of ~370 bar^{37,38}. Studies are in progress of the mineralogy, composition and isotopic characteristics of the various hydrothermal mineral phases in the recovered samples (G. Thompson and P.A.R., unpublished data).

Conclusions

The discovery of black smokers, chimneys, massive sulphides and vent biota fills a major gap in the understanding of slow-spreading oceanic ridges. The low-temperature hydrothermal zone previously delineated at mid-depth on the east wall of the rift valley at the TAG hydrothermal field is not an isolated phenomenon, but is part of a larger hydrothermal system, probably related by physical and chemical gradients to the high-temperature area at the juncture between the wall and the floor of the rift valley (Fig. 1). The high-temperature area, with vigorous venting of black smokers and precipitation of massive sulphides, is laterally contiguous with the low-temperature zone, with weak venting of warm springs and precipitation of oxides, hydroxides and silicates.

The elliptical configuration of the compound mound, with black smokers venting from the inner mound and evidence of

lava flows and extensive fracturing at the perimeter of the outer mound, suggests that the mound may overlie a crater partially collapsed over an off-axis magma chamber (Fig. 2). The parallelism between the long axis of the ellipse and the axis of the rift valley (020°) suggests the additional influence of faulting sub-parallel to that axis. The observation that the black smokers vent from fractures rather than from chimneys suggests that the fractures opened in response to an episode of seismicity so recent that chimneys have not had time to grow. The massive sulphides recovered and the inactive chimneys imaged formed during previous episodes of black smoker activity. The relatively slow observed rate of ascent of the plumes indicates either hydrothermal fluid temperatures lower than that of Pacific black smokers or, alternatively, high temperatures with thermal expansion partially compensated by the observed dense particle loading and/or phase separation³⁹. Fluid temperature measurements and sampling by submersible are needed to resolve this question. Assuming that the inner mound is constructed primarily of hydrothermal precipitates (average density 3.0 g cm^{-3}) in the form of a semi-ellipsoid, the deposit would weigh 4.5×10^6 tonnes. The similarities in size, shape and structure of this mineral deposit to certain ancient massive sulphide deposits in the form of convex-upward mounds, such as Noranda-type deposits⁴⁰, is striking, although differences exist in petrological associations.

The record of hydrothermal activity at the TAG hydrothermal field is partially preserved in localized ponded sediment. Fluctuations in activity through time may be distinguished from spatial variations within the field by examination of this sedimentary record^{1,16,29}. The relation between inactive chimneys (Figs 1, 4) and black smokers (Fig. 3) may also be deciphered from the sedimentary record. The occurrence of the full range of seafloor hydrothermal phenomena at this site on the Mid-Atlantic Ridge suggests a role for slow-spreading oceanic ridges in the hydrothermal exchange processes that affect ocean chemistry, seafloor mineralization, heat transfer and biological adaptation.

We thank E. Bernard, director, and S. Hammond, manager of the NOAA Vents program, for encouragement and support; H. Bezdek and D. K. Atwood of NOAA, AOML for support; A. Meyer, S. Jones and G. Eberhart for work with the camera/temperature system and assistance in data processing; acting commanding officer C. Molyneux, operations officer K. A. Doggett, other officers and crew of the NOAA ship *Researcher* for capable support at sea; the NOAA Vents program and NERC (UK) for funding; L. Burk (Oceano Instruments) for acoustic navigation; J. R. Delaney (University of Washington) for discussion of pressure-temperature-density effects on hydrothermal solutions; B. W. Haynes for sample analysis as part of the Characterization of Seafloor Minerals project of the US Bureau of Mines; and J. F. Grassle (Woods Hole Oceanographic Institution) for guidance of biological studies.

Received 15 January; accepted 11 March 1986.

1. Rona, P. A. *et al.* *J. geophys. Res.* **89**, 11365–11377 (1984).
2. Francheteau, J. *et al.* *Nature* **277**, 523–528 (1979).
3. Spiess, E. N. *et al.* *Science* **207**, 1421–1444 (1980).
4. Hekinian, R., Frevier, M., Bischoff, J. L., Picot, P. & Shanks, W. C. *Science* **207**, 1433–1444 (1980).
5. Lonsdale, P. F., Bischoff, J. L., Burns, V. M., Kastner, M. & Sweeney, R. E. *Earth planet. Sci. Lett.* **49**, 8–20 (1980).
6. Normark, W. R. *et al.* *Mar. Technol. Soc. J.* **16**, 46–53 (1982).
7. Chase, R. L. *et al.* *Nature* **313**, 212–214 (1985).
8. Paull, C. K. *et al.* *Science* **226**, 965–967 (1984).
9. Kennicutt, M. C. II *et al.* *Nature* **317**, 351–353 (1985).
10. Kong, L. *et al.* *Eos* **66**, 936 (1985).
11. Klinkhammer, G., Rona, P., Elderfield, H. & Greaves, M. *Eos* **66**, 936 (1985).
12. Nelsen, T. A., Klinkhammer, G. & Trefry, J. *Eos* **66**, 936 (1985).
13. Rona, P. A. *Eos* **66**, 682–683 (1985).
14. Rona, P. A. *Eos* **66**, 936 (1985).
15. Trefry, J. E. & Trocine, R. P. *Eos* **66**, 936 (1985).
16. Metz, S. & Trefry, J. H. *Eos* **66**, 936 (1985).
17. Saunders, P. M. & Fofonoff, N. P. *Deep Sea Res.* **23**, 109–111 (1976).
18. Rona, P. A. *Ocean Mgmt* **1**, 145–159 (1973).
19. Rona, P. A. & Scott, R. B. (convenors) *Eos* **55**, 292–295 (1974).
20. Scott, R. B., Rona, P. A., McGregor, B. A. & Scott, M. R. *Nature* **251**, 301–302 (1974).

21. Rona, P. A. *J. geol. Soc. Lond.* **137**, 385–402 (1980).
22. Rona, P. A., Harbison, R. N., Bassinger, B. G., Scott, R. B. & Nalwalk, A. J. *Geol. Soc. Am. Bull.* **87**, 661–674 (1976).
23. Lattimore, R. K., Rona, P. A. & DeWald, O. J. *geophys. Res.* **79**, 1207–1209 (1974).
24. McGregor, B. A. & Rona, P. A. *J. geophys. Res.* **80**, 3307–3314 (1975).
25. Temple, D. G., Scott, R. B. & Rona, P. A. *J. geophys. Res.* **84**, 7453–7466 (1979).
26. Karson, J. A. & Rona, P. A. *Eos* **63**, 1103 (1982).
27. Thompson, G., Mottl, M. J. & Rona, P. A. *Chem. Geol.* **49**, 243–257 (1985).
28. Jenkins, W. J., Rona, P. A. & Edmonds, J. M. *Earth planet. Sci. Lett.* **49**, 39–44 (1980).
29. Shearme, S., Cronan, D. S. & Rona, P. A. *Mar. Geol.* **51**, 269–291 (1983).
30. Francheteau, J., Juteau, T. & Rangan, C. *Nature* **281**, 209–211 (1979).
31. Corliss, J. B. *et al.* *Science* **203**, 1073–1083 (1979).
32. Edmond, J. M. *Eos* **66**, 918 (1985).
33. Macdonald, K. C., Becker, K., Spiess, F. N. & Ballard, R. D. *Earth planet. Sci. Lett.* **48**, 1–7 (1980).
34. Converse, D. R., Holland, H. D. & Edmond, J. M. *Earth planet. Sci. Lett.* **69**, 159–175 (1984).
35. Weiss, R. F., Lonsdale, P., Lupton, J. E., Bainbridge, A. E. & Craig, H. *Nature* **267**, 600–603 (1977).
36. Delaney, J. R., McDuff, R. E. & Lupton, J. E. *Eos* **65**, 973 (1984).
37. Blount, C. W. & Dickson, F. W. *Geochim. cosmochim. Acta* **33**, 227–245 (1969).
38. Janecky, D. R. & Seyfried, W. E. *Jr Am. J. Sci.* **283**, 831–860 (1983).
39. Delaney, J. R. & Cosens, B. *Mar. Technol. Soc. J.* **16**, 62–66 (1982).
40. Franklin, J. M., Lydon, J. W. & Sangster, D. F. *Econ. Geol.* **75th Anniv. Vol.** 485–627 (1981).
41. Williams, A. B. & Rona, P. A. *J. crustacean Biol.* (in the press).

Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*

M. A. Anderson*, E. C. Cornish*, S.-L. Mau*, E. G. Williams*, R. Hoggart*, A. Atkinson*, I. Bonig†, B. Grego†, R. Simpson†, P. J. Roche‡§, J. D. Haley‡§, J. D. Penschow‡§, H. D. Niall‡§, G. W. Tregear‡§, J. P. Coghlan‡§, R. J. Crawford‡§ & A. E. Clarke*

* Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville, Victoria 3052, Australia

† Joint Protein Structure Laboratory, Ludwig Institute for Cancer Research (Melbourne Branch) and The Walter and Eliza Hall Institute for Medical Research, Parkville, Victoria 3052, Australia

‡ Xerogen Pty Ltd, c/o 121 William Street, Melbourne, Victoria 3052, Australia

Self-incompatibility in flowering plants is controlled by the S gene. A complementary DNA clone encoding a style protein of Nicotiana alata which segregates with the S₂ allele has now been sequenced. The S-allele-associated style components in different genotypes of N. alata and in Lycopersicon peruvianum, another self-incompatible species in the family Solanaceae, are homologous.

OVER half the flowering plant families have specific self-incompatibility genes (*S* genes) which prevent inbreeding (for reviews see refs 1–6). Among homomorphic multi-allelic incompatibility systems there are two major types, showing either gametophytic or sporophytic control of the self-incompatibility response. Gametophytic incompatibility is the more common and, in many cases, is controlled by a single nuclear gene locus (*S* locus) with multiple alleles. Pollen expresses its haploid *S* genotype and matings are incompatible if the *S* allele of the pollen is matched by one of two *S* alleles expressed in the diploid tissue of the pistil. During both incompatible and compatible matings, pollen germinates and the pollen tubes grow through the stigma into the transmitting tissue of the style. Tube growth from incompatible pollen grains is arrested in the style; the tip of the tube becomes thickened, there is often a characteristic deposit of callose close to the tip, and the tip may burst. In sporophytic incompatibility, on the other hand, pollen behaviour is determined by the genotype of the pollen-producing plant; inhibition usually occurs on the stigma surface and is often associated with deposition of callose in the stigma cells. The gametophytic system is generally considered to be the ancestral form of self-incompatibility in flowering plants, with the sporophytic system being derived from it¹.

The genetics of gametophytic self-incompatibility systems has been studied extensively (see, for example, refs 1, 7–11). Several attempts have been made to identify the active product(s) of the *S* gene in both gametophytic and sporophytic systems^{12–14}, and recently a clone encoding part of an *S*-locus-specific glycoprotein from *Brassica oleracea* stigmas has been described¹⁵. In general, the early studies point to the presence of *S*-genotype-associated antigens or proteins in the pistil, but have so far failed to positively identify corresponding molecules in pollen or pollen tubes. Although several genetic and molecular models have been constructed to account for the breeding behaviour of *S* alleles^{16–18}, the detailed nature of the *S*-gene product and its mode of action in preventing fertilization are still unknown.

The major criteria for positive identification of an *S*-allelic product are, first, that its presence in reproductive tissues of the plant must consistently be associated with the corresponding breeding behaviour and, second, that it should show specific biological activity under appropriate bioassay conditions (for example, it should inhibit growth of pollen tubes bearing the same allele but not tubes bearing a different allele).

To elucidate the molecular basis of incompatibility, we have isolated an *S*-genotype-associated glycoprotein from styles of an ornamental tobacco, *Nicotiana alata* Link and Otto, a species which displays gametophytic self-incompatibility. Here we report the association of this glycoprotein with expression of the *S*₂ allele and describe a complementary DNA clone encoding the protein. We have used this cDNA to examine both expression of self-incompatibility in female tissues of the flower and the homology between different *S* alleles of *N. alata*.

The present study represents an initial step in understanding how plants discriminate between self and non-self¹⁹. We examine how 'self' pollen tubes are recognized and how their growth is arrested within the female tissues of the style. This system is defined in terms of classical genetics by breeding techniques, and we have now started to define it in terms of molecular genetics.

Expression of *S*₂ allele

From the original acquisition, *S*₂*S*₃, *S*₂*S*₂ and *S*₃*S*₃ homozygotes were produced by bud self-pollination. Bud self-pollination of a second acquisition, *S*₁*S*₃, provided the homozygote *S*₁*S*₁, and a third heterozygote, *S*₁*S*₂, was obtained by intercrossing the two acquisitions. Extracts from styles of mature plants revealed an association of a component of relative molecular mass (*M*_r) 32,000 (32K) with the presence of the *S*₂ allele; this was apparent on single-dimension SDS-polyacrylamide gels (not shown) and in two-dimensional gels (Fig. 1A), which showed that the 32K component had a high *pI*. The high-*pI* proteins in extracts of styles bearing different *S* alleles could be compared on the same isoelectric focusing gel by selecting from the gel those components in the high-*pI* range (>9.5) and subjecting them to SDS-polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension²⁰. Association of the 32 K band with *S*₂ allele was clearly demonstrated in this selective two-dimensional gel electrophoresis (Fig. 1B).

Style proteins were analysed for 53 plants resulting from reciprocal crosses involving the genotypes *S*₂*S*₂, *S*₃*S*₃, *S*₂*S*₃ and *S*₁*S*₃, using conventional two-dimensional gel electrophoresis and selective two-dimensional gel electrophoresis. The results demonstrated genetic segregation of the 32K component with breeding behaviour identifying the *S*₂ allele. Forty progeny plants in which the 32K component was present in style extracts, all rejected *S*₂ pollen. The remaining 13 plants, apparently lacking the 32K component, all accepted *S*₂ pollen. Thus, no recombination was observed between *S*₂ breeding behaviour and presence of the 32K component in the style.

§ Present addresses: Howard Florey Institute of Experimental Physiology and Medicine, University of Melbourne, Parkville, Victoria 3052, Australia (P.J.R., J.D.P., G.W.T., J.P.C., R.J.C.); Imperial Cancer Research Fund, London WC2A 3PX, UK (J.D.H.); Genentech Inc., South San Francisco, California 94080, USA (H.D.N.).

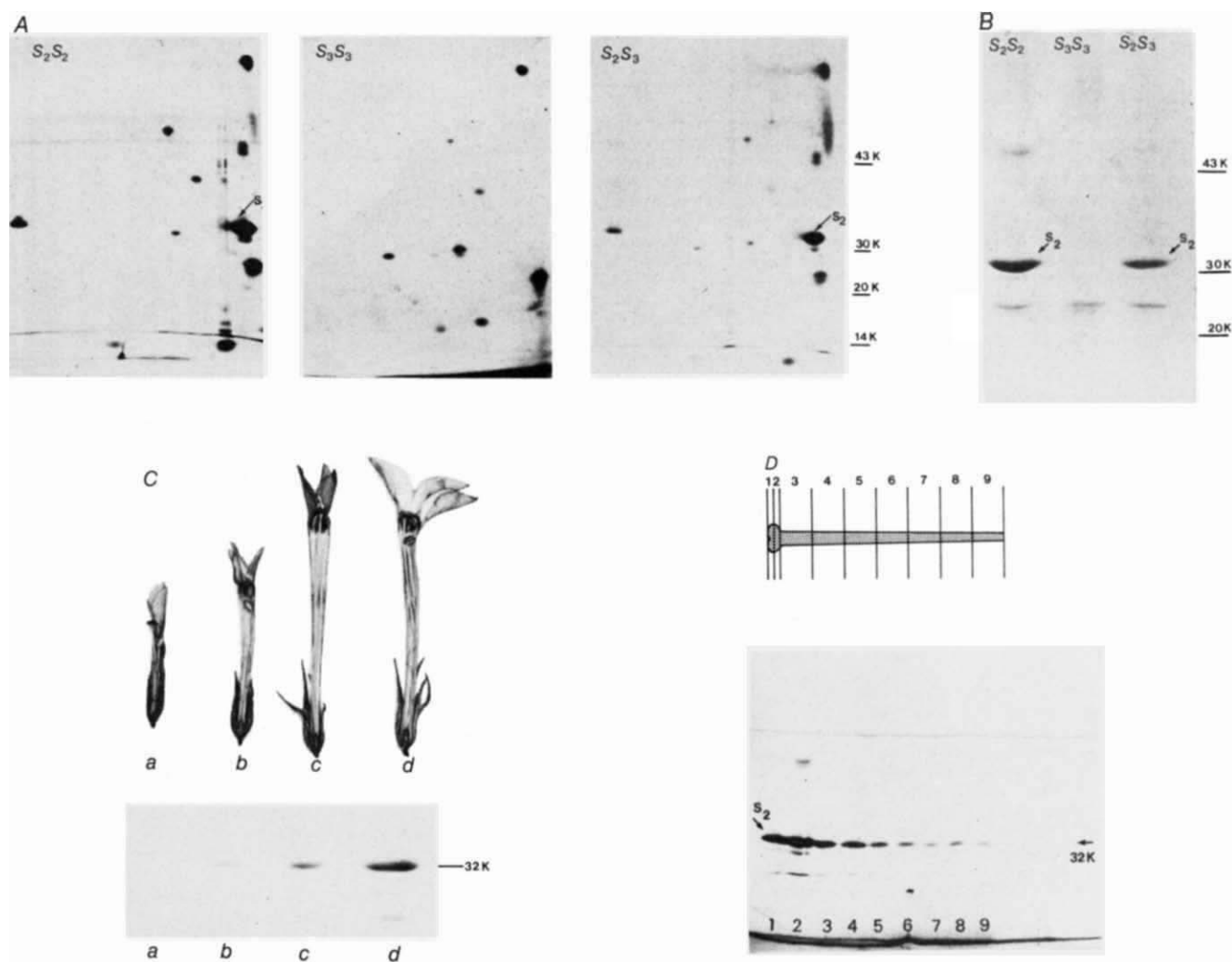


Fig. 1 Correlation of expression of the S_2 allele of *Nicotiana alata* with the presence of a 32K glycoprotein in style extracts. **A**, Two-dimensional gel electrophoresis of style extracts from mature plants of *N. alata*, genotypes S_2S_2 , S_3S_3 and S_2S_3 . A high- pI component of M_r 32,000 is present in the parent acquisition S_2S_3 and in bud-selfed progeny with the genotype S_2S_2 , assigned by test crossing. The 32K component is absent from progeny of the same bud-selfing with an assigned genotype of S_3S_3 . **B**, Selective two-dimensional gel electrophoresis of style extracts from mature plants as shown in **A**. The 32K band is present in style extracts of all plants which carry a functional S_2 allele. **C**, Selective two-dimensional gel electrophoresis of style extracts from flowers of genotype S_2S_3 at different developmental stages. The 32K band is absent from style extracts prepared at the green-bud stage when the self-incompatibility response is absent (**a**), but appears, together with expression of self-incompatibility, at the time of petal colouration (**b**) and is present in higher concentrations in extracts prepared from styles of mature flowers (**c**, **d**). **D**, Selective two-dimensional gel electrophoresis of extracts from segments of mature styles. A diagram representing the stigma and upper part of the style is shown; the stigma sections were 1 mm long and the style sections 6 mm long. The 32K glycoprotein is present at highest concentrations in the sections which include the stigma (1 and 2), and in the upper segments of the style (3, 4 and 5).

Methods. **A**, Seed of *N. alata* of incompatibility genotypes S_2S_3 and S_1S_3 was provided by Dr K. K. Pandey. Crosses were made by hand on emasculated flowers in a pollinator-proof glasshouse and scored as incompatible if flowers abscised without capsule development, or compatible if filled seed capsules were formed. Plants homozygous for the S_2 and S_3 alleles were obtained by self-pollination of immature flowers at the green-bud stage using mature pollen from the same plant and stigmatic secretion from a mature, emasculated flower to aid pollen germination on the surface of the immature stigma. As incompatibility is first expressed at the petal colouration stage, self-pollen tubes are not always arrested when flowers are pollinated at the green-bud stage. Abscission of fruits containing low numbers of 'selfed' seeds was prevented by smearing the floral pedicels with 1% indoleacetic acid (IAA) in crude lanolin. Genotypes were assigned by test crossing to original acquisitions. Style extracts were prepared by crushing six styles lightly in 0.3 ml of extraction buffer [50 mM Tris-HCl pH 8.5, 10 mM EDTA, 1 mM CaCl_2 , 1 mM phenylmethylsulphonyl fluoride (PMSF), 1 mM dithiothreitol (DTT)]. The extracts were centrifuged and the protein concentrations measured³⁰. Extracts containing 15 μg of protein were analysed on two-dimensional gels run according to Mau *et al.*²⁰. Protein spots were visualized by silver staining. **B**, Extracts were prepared from six mature styles as described for **A**, and samples containing 15 μg of protein were loaded onto the isoelectric focusing gel. After running according to Mau *et al.*²⁰, the region corresponding to the pI range >9.5 , which contains the S_2 -glycoprotein, was excised and run on a 15% SDS-polyacrylamide gel and stained with Coomassie blue. As many of the proteins present in the extract are of low pI , the pattern shown in conventional two-dimensional gels is considerably simplified by this procedure. **C**, Style extracts of flowers at the developmental stages shown were prepared and examined by selected two-dimensional gel electrophoresis as described for **B**. **D**, Six mature styles were harvested, aligned and sectioned with a razor blade according to the diagram. Each group of sections was extracted separately with 0.3 ml of buffer and the extracts examined by selective two-dimensional gel electrophoresis as described for **B**. Protein bands were visualized by silver staining.

Selective two-dimensional gel electrophoresis was also used to demonstrate the association of the 32K component with development of S_2 function in maturing styles (Fig. 1C). The 32K component was present at highest concentrations in upper segments of the style (Fig. 1D), the zone in which the arrest of incompatible pollen tubes occurs. Thus, the 32K glycoprotein segregates with S_2 breeding behaviour, appears concomitantly

with S_2 function in maturing styles and occurs in highest concentration in the part of the style in which inhibition of incompatible pollen tube growth occurs. This 32K glycoprotein is termed the S_2 -glycoprotein. The appearance of S -allele-associated style proteins in mature but not immature styles has been noted for other self-incompatible species, and the concentration of S -allele-associated proteins in the stigma and upper style

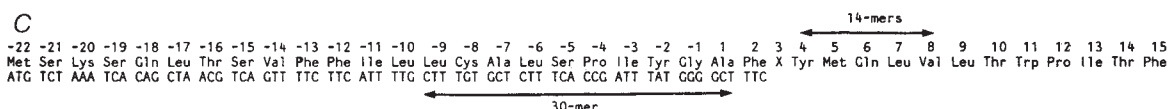
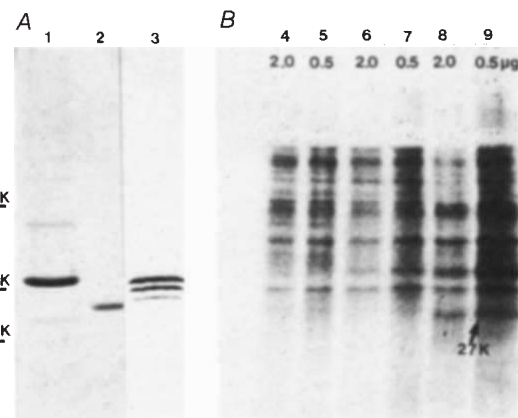


Fig. 2 Protein products specific to the S_2 allele. **A**, Purification and deglycosylation of the S_2 -glycoprotein. Lane 1, S_2 -glycoprotein isolated by chromatographic procedures (M_r 32,000); lane 2, S_2 -glycoprotein after chemical deglycosylation (M_r 26,000); lane 3, S_2 -glycoprotein after enzymatic deglycosylation. Four species are detected by silver staining: a strongly staining band of 32K which corresponds to the undegraded glycoprotein, two 30K species running closely together (the separation of these is not clear in the photograph) and a fourth species of M_r 28,000. **B**, Identification of an *in vitro* translation product specific to mature styles. The figure shows *in vitro* translation products of mRNA isolated from *N. alata* ovary (lanes 4 and 5), green-bud style (lanes 6 and 7) and mature style (lanes 8 and 9). The amount of mRNA used in each experiment is indicated. The 27K translation product unique to mature style mRNA is indicated. **C**, Amino-acid sequence of the N-terminus of the mature S_2 -glycoprotein and the C-terminus of a putative signal sequence for this protein, as determined by N-terminal analysis and deduced from DNA sequence analysis. Amino acids 1–15 were determined by N-terminal analysis of the isolated glycoprotein where X denotes no clear assignment. From the sequence Tyr-Met-Gln-Leu-Val (amino acids 4–8), the following three pools of oligodeoxyribonucleotide 14-mers, each of 8-fold degeneracy, were prepared: 5' d(ACA^GAGT^CTGCATA^GTA (pool I), 5' d(ACC^TAGT^CTGCATA^GTA (pool II) and 5' d(ACC^TAAT^CTGCATA^GTA (pool III). The nucleotide sequence complementary to a portion of cDNA synthesized when pool I was used to prime cDNA synthesis is shown. This sequence is in the same sense as the mRNA and encodes 22 amino acids of a putative signal peptide and two amino acids at the N-terminus of the mature S_2 -glycoprotein. A synthetic 30-mer probe complementary to the sequence indicated was synthesized as described below.

Methods. **A**, Flowers were emasculated at the onset of petal colouration; 2 days later, the fully mature styles were removed and stored at -70°C . (Styles refer to stigma and style which were removed together; the ovary was not included.) Frozen styles (3 g; 220 styles) were ground to a fine powder in liquid nitrogen using a mortar and pestle, then ground further in 50 ml of extraction buffer [0.1 M Tris-HCl pH 8.5, 1 mM CaCl_2 , 0.1 M NaCl, 1 mM PMSF, 1 mM DTT, 10 mM EDTA and 1% (w/w) insoluble polyvinyl pyrrolidone (Polyclar AT)]. The homogenate was centrifuged (10,000g; 15 min) and the supernatant equilibrated with 5 mM ammonium bicarbonate pH 8.6, 1 mM NaCl, 1 mM CaCl_2 , 1 mM EDTA, by passage through a Sephadex G-25 column (1.6 × 22 cm; void volume 11 ml). The first 16 ml that eluted after the void volume was collected (28.6 mg total protein as bovine serum albumin) and applied to DEAE-Sephacrose (2 × 13 cm), equilibrated with the same ammonium bicarbonate buffer. As the S_2 -associated glycoprotein is known from isoelectric focusing experiments to have a high pI (>9.5), it does not bind to DEAE-Sephacrose at pH 8.6, but was present in the unbound fraction together with one other major component and several minor components. The unbound fractions were combined (2.78 mg protein as BSA), concentrated to 16 ml and fractionated further by chromatography on concanavalin A (Con A)-Sephacrose. Then the Con A-Sephacrose was washed with 5 vol. 0.1 M methyl- α -D-mannoside in 10 mM sodium acetate pH 7.0, 0.1 M NaCl, 1 mM MgCl_2 , 1 mM CaCl_2 , 1 mM MnCl_2 and subsequently with sodium bicarbonate (0.25 M, pH 8.8) (four changes during a 1-h period). Four volumes of sodium bicarbonate (0.25 M pH 8.8) containing 0.03% glutaraldehyde were added to the beads, the two were mixed and then the beads were washed with 0.1 M sodium bicarbonate pH 8.0, containing 0.5 M NaCl, resuspended in the sodium acetate buffer (without methyl- α -D-mannoside) and packed into a column (0.8 × 14 cm). The unbound fraction from DEAE-Sephacrose was equilibrated in acetate buffer by passage through a Sephadex G-25 column equilibrated with acetate buffer, then applied to the Con A-Sephacrose column. Unbound material was collected, the column was washed with 10 vol. acetate buffer, and the bound material eluted with 0.2 M methyl- α -D-mannoside in acetate buffer. The eluted material was separated from the methyl- α -D-mannoside by passage through a Biogel PI50 column. Yield was 490 μg of protein (1.7% of total style protein). Chemical deglycosylation of the S_2 -associated glycoprotein: the isolated S_2 -glycoprotein (lane 1) was deglycosylated (lane 2) using the trifluoromethanesulphonic acid (TFMS) procedure³¹, and the deglycosylated proteins were analysed on 15% SDS-polyacrylamide gels which were stained with Coomassie blue. Enzymatic deglycosylation of the S_2 -associated glycoprotein: Isolated glycoprotein (5 μg) in SDS sample buffer (10 μl) was heated at 100°C for 2 min, then cooled and 46.5 U endoglycosidase F (Miles Laboratories) added. The mixture was incubated at 35°C for 2 h, then applied to a 15% SDS-polyacrylamide gel, which was stained with silver stain. **B**, Poly(A)⁺ mRNA (0.5 or 2.0 μg) was isolated from mature styles, green-bud styles or ovary³², and translated using an amino-acid-depleted rabbit reticulocyte lysate kit (Amersham) in the presence of tritiated leucine, lysine, phenylalanine, proline and tyrosine. Translation products were separated by SDS-PAGE using 12.5–20% SDS-polyacrylamide gradient gels. The products were visualized by autoradiography after treatment of the gel with Amplify (Amersham). **C**, N-terminal sequencing of the S_2 -glycoprotein purified as described in **A** was performed using an Applied Biosystems Model 470 A gas-phase sequencer. Isolated glycoprotein (~200 μg) was applied in aqueous solution to a glass fibre disk and evaporated to dryness. The resultant phenylthiohydantoin amino-acid derivatives were identified on an IBM-CN column (IBM, Danbury, Connecticut) using a sodium acetate-acetonitrile trifluoroacetic acid gradient, in a Beckman-Altex high-pressure liquid chromatograph. From the N-terminal sequence, three pools of 14-mers (oligodeoxyribonucleotides) of 24-fold degeneracy to Tyr-Met-Gln-Leu-Val were synthesized by the solid-phase phosphoramidite method³³ using an Applied Biosystems (ABI) Model 380A DNA Synthesizer. The 14-mers were labelled at the 5' end using [α -³²P]ATP (5,000 Ci mmol⁻¹) and T₄ polynucleotide kinase. The labelled 14-mers (500 μg ml⁻¹) were used to prime cDNA synthesis. Specifically primed cDNA products were separated on an 8% acrylamide/7 M urea gel. Oligonucleotide pool I primed the synthesis of a cDNA fragment of ~100 nucleotides, whereas no cDNA products were detected when either pool II or pool III was used to prime cDNA synthesis. The cDNA fragment was purified and sequenced³⁴ to give unequivocal assignment for 72 nucleotides as shown. From this sequence, a single 30-mer (oligodeoxyribonucleotide) complementary to the cDNA sequence encoding amino acids -9 to +1 was synthesized³³. This 30-mer was used to screen a cDNA library as described in Fig. 3 legend.



segments has been noted for *Lycopersicon peruvianum*²⁰, another member of the Solanaceae that has a gametophytic self-incompatibility system. Previous studies with *N. alata*²¹ also demonstrated the developmental regulation and distribution of S -allele-associated style components.

The absolute correlation between the presence of the 32K glycoprotein component in the style and the ability of the style to reject S_2 pollen, represented strong evidence that the 32K protein was a product of either the S_2 allele or of a gene closely linked to it that does not appear to recombine with S_2 in the present breeding programme. We attempted to isolate the S_2 product and to clone the cDNA encoding the protein.

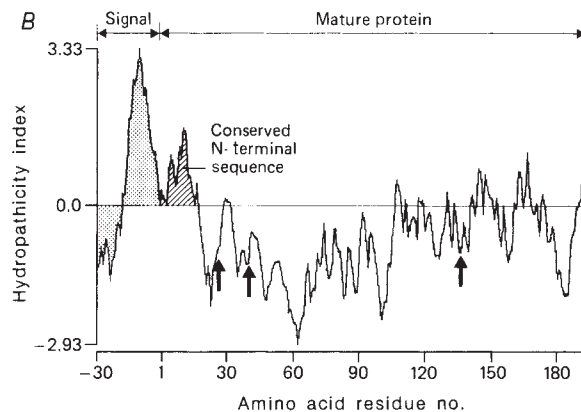
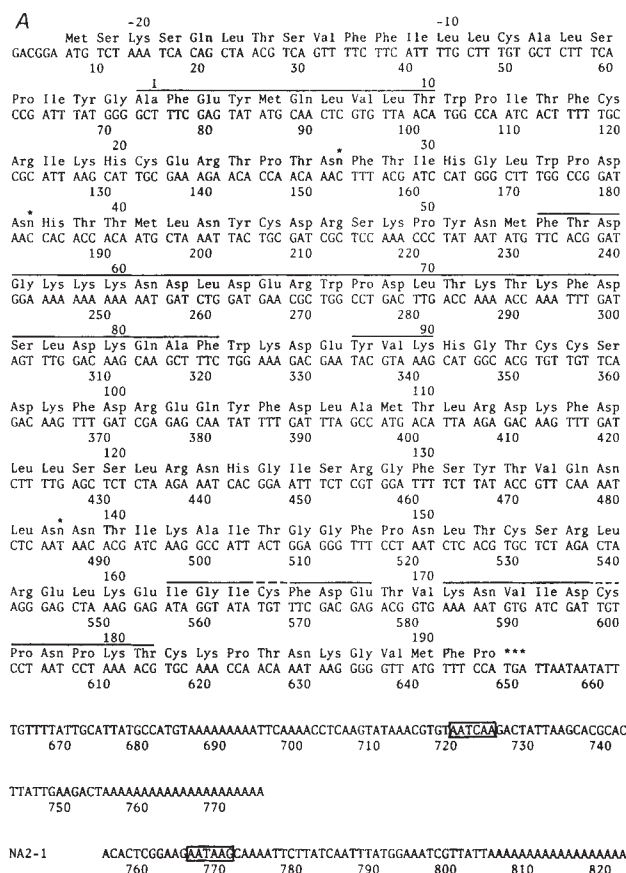
Isolation of S_2 -glycoprotein cDNA

A method for isolating the 32K component from styles was established, based on the fact that it is a glycoprotein with a

high pI . Chemical deglycosylation of the 32K component, isolated by a combination of ion-exchange and affinity chromatography, yielded a single product (M_r 26K) (Fig. 2A). The carbohydrate component contains fucose, arabinose, xylose, mannose, galactose and possibly glucose, but the organization of these monosaccharide residues into carbohydrate chains is unknown. At least part of the carbohydrate occurs as N -linked chains, as the isolated component was susceptible to degradation with endoglycosidase F (Endo F), yielding three products as well as the undegraded glycoprotein of M_r 32K (Fig. 2A). The Endo F products comprised two species running closely together, each with an estimated M_r of 30,000, and a single species of M_r 28,000. A product of M_r 27,000, which is slightly larger than the chemically deglycosylated protein, was found on *in vitro* translation of messenger RNA isolated from a mature style but was absent from translation products of

Fig. 3 A, The nucleotide sequence and predicted amino-acid sequence of the cDNA encoding the S_2 -glycoprotein of *N. alata*, as defined by the sequence of clone NA-2-2. Nucleotides 19–754 are identical to those of clone NA-2-1. The cDNA insert from clone NA-2-1 had an additional 50 nucleotides at the 3'-untranslated end preceding a polyadenylation site, as shown beneath the last line of the NA-2-2 sequence. Only one strand with the polarity of the mRNA is shown. Nucleotides are numbered beneath the DNA sequence. The predicted amino-acid sequence is given above the nucleotide sequence and is numbered beginning with 1 for the first amino acid of the mature protein. The putative signal peptide is indicated by negative numbers. The overlined amino acids are those residues identified by protein sequencing. Broken overlining indicates those residues that could not be identified unambiguously by amino-acid sequencing. Potential glycosylation sites are indicated by asterisks, and potential polyadenylation signals at residues 720–725 and 766–771 (for NA-2-1) are boxed. B, A hydrophobicity plot of the deduced amino-acid sequence of the unprocessed S_2 -glycoprotein. The region of the profile corresponding to the putative signal peptide (amino acids –22 to –1) is stippled. The hydrophobic N-terminal sequence, which exhibits striking homology to the S-allele-associated style components from *L. peruvianum*, is shaded. The positions of the three putative N-glycosylation sites are indicated by arrows.

Methods. A, Poly(A)⁺ mRNA isolated from mature S_2S_3 styles³² was used to construct a cDNA library in λ gt10 (ref. 35). The library was screened for sequences specific to mature style by differential hybridization to single-stranded ³²P-cDNAs prepared from ovary, green-bud style and mature style poly(A)⁺ mRNAs. Prehybridization (2 h at 42 °C) was in 5 × Denhardt's, 5 × SSC, 50 μ g ml^{–1} sonicated salmon sperm DNA, 50 mM Na-phosphate pH 6.8, 1 mM Na-pyrophosphate, 100 μ M ATP and 50% deionized formamide. The cDNA probes were used at 4×10^3 c.p.m. ml^{–1}. Filters were washed in 0.1 × SSC, 0.1% SDS at 42 °C. One clone (NA-2-1) which hybridized to mature style cDNA, but not to cDNA from ovary or green-bud style, also hybridized to the synthetic 30-mer described in the legend to Fig. 2C. The same hybridization solution as described above was used for hybridizations with the 30-mer (10^8 c.p.m. μ g^{–1}, 0.4 – 1×10^6 c.p.m. ml^{–1}), except that the formamide concentration was 20% and the temperature was decreased to 37 °C. A second cDNA library in λ gt10 was constructed by a method designed to optimize isolation of full-length clones²². A library containing 200,000 plaques was obtained from 5 μ g of mature style poly(A)⁺ RNA. From this library, a clone, designated NA-2-2, which hybridized to the *Eco*RI insert from NA-2-1 and the synthetic 30-mer, was identified. The insert was labelled with [α -³²P]dATP using DNA polymerase I (Klenow fragment)³⁶. The 30-mer was labelled with [γ -³²P]ATP using T₄ polynucleotide kinase. The cDNA inserts from clones NA-2-1 and NA-2-2 were ligated into the bacteriophage vector M13mp8 RF DNA³⁷ and sequenced by the chain termination method^{38,39}. Sequence data obtained in this way were analysed and manipulated using the computer programs of Staden⁴⁰, as modified by Dr A. P. Kyne. Peptide sequencing: The 32K glycoprotein was electroeluted from selective two-dimensional SDS-polyacrylamide gels (60 μ g) and purified by 'reverse-gradient' HPLC²⁰ before reduction and alkylation⁴¹ and desalting by 'reverse-gradient' HPLC²⁰. The sample was divided into two equal portions and fragmented by either CNBr cleavage⁴² or *S. aureus* protease digestion using an enzyme/substrate ratio of 1:50 in 1% ammonium bicarbonate⁴³. The peptides were purified for sequence analysis by multi-dimensional microbore HPLC⁴¹. HPLC was performed on a Perkin-Elmer liquid chromatograph (model LC4) equipped with a model LC95 ultraviolet detector, a model LS4 fluorimeter and a model LS1 100 integrator. Fractions were collected either with a Pharmacia Frac 100 or manually. B, The hydrophobicity profile was determined using the predictive rules of Kyte and Doolittle⁴⁴ and a span of nine consecutive residues.



mRNA from both ovary and styles harvested at the green-bud stage (Fig. 2B). As the 32K S_2 -glycoprotein was not detected in extracts of ovary or styles from flowers at the green-bud stage, it was concluded that the 27K component corresponds to the *in vitro*-translated product of S_2 mRNA. This component is slightly larger than the 26K chemically deglycosylated glycoprotein, consistent with the presence of a signal peptide that is cleaved during secretion of the S_2 -glycoprotein from the transmitting tract cells to the extracellular space.

The N-terminal sequence of the isolated glycoprotein was determined and three pools of 14-mers of 24-fold degeneracy to amino acids 4–8 were prepared and used to prime cDNA synthesis (see Fig. 2 legend). A specifically primed fragment of ~100 nucleotides, produced using one of the three pools of 14-mers, was purified and sequenced. From this sequence a single 30-mer complementary to the cDNA sequence encoding amino acids –9 to +1 was synthesized (Fig. 2C). A cDNA library was prepared in λ gt10 using mRNA isolated from mature

styles of *N. alata* of genotype S_2S_3 . The library was subjected to differential screening using ³²P-labelled cDNA prepared from mRNA isolated from mature and green-bud styles. Plaques that hybridized strongly only to the mature style cDNA were selected and rescreened using cDNA prepared from mature style and ovary mRNA. Ovary cDNA was chosen for the second screen because electrophoretic analysis of extracts of mature style and ovary indicated that these tissues have in common some proteins that are not present in green-bud styles (data not shown). The protein profiles of other plant organs were too diverse compared with mature style to be useful for differential screening. Clones that hybridized to mature style cDNA, but not to ovary or green-bud style cDNA, were screened with the synthetic oligodeoxynucleotide 30-mer described above. One of the clones that hybridized with the 30-mer as well as the cDNA prepared from mature style mRNA, but not ovary or green-bud style mRNA, was chosen for further study. This clone was designated NA-2-1.

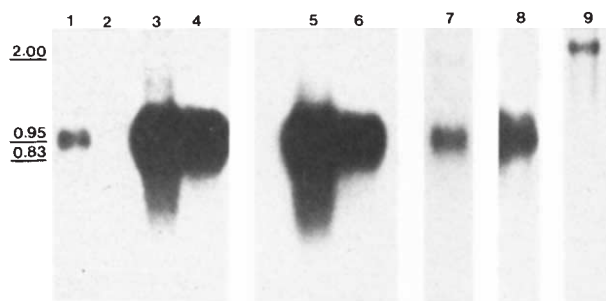


Fig. 4 Northern blot analysis of mRNAs from styles of different *S* genotypes isolated from *N. alata* and *L. peruvianum*. Lanes 1–4, *N. alata* genotypes S_2S_3 (lane 1, ovary 5 μ g; lane 2, green bud 5 μ g; lane 3, mature style 2 μ g; lane 4, mature style 0.5 μ g); lanes 5, 6, *N. alata* genotype S_2S_2 (lane 5, mature style 5 μ g; lane 6, mature style 0.5 μ g); lane 7, *N. alata* genotype S_3S_3 (mature style 5 μ g); lane 8, *N. alata* genotype S_1S_3 (mature style 5 μ g); lane 9, *L. peruvianum* mature style (5 μ g). Numbers on the left are size markers.

Methods. mRNAs were isolated as described previously³² and electrophoresed in formaldehyde/1.2% agarose gels⁴⁵ using a running buffer containing 20 mM morpholinopropanesulphonic acid pH 7.0, 5 mM sodium acetate and 0.1 mM EDTA pH 8.0. The RNAs were transferred to nitrocellulose filters and probed with ³²P-labelled S_2 cDNA (10^8 c.p.m. μ g⁻¹, 2×10^6 c.p.m. ml⁻¹). The size markers used were *EcoRI*/*HindIII* restriction fragments of λ DNA. Prehybridization (2 h at 42 °C) and hybridization (16 h at 42 °C) was carried out in 5 $\times$ Denhardt's, 5 $\times$ SSC, 50 μ g ml⁻¹ sonicated salmon sperm DNA, 100 μ M ATP and 50% deionized formamide. Filters were washed in 0.1 $\times$ SSC, 0.1% SDS at 42 °C. All exposures were for 3 h, except for S_1S_3 *N. alata* mRNA (lane 8) and S_1S_3 *L. peruvianum* mRNA, where films were exposed for 18 h.

Nucleotide sequence analysis of the cDNA insert in clone NA-2-1 showed that it contained an open reading frame, but had no ATG that could act as an initiation codon. To obtain a clone containing a full-length cDNA insert, another cDNA library was prepared using a method²² which optimizes the recovery of full-length clones. This library (200,000 plaques) was screened with both the cDNA insert from NA-2-1 and the 30-mer shown in Fig. 2C. A second clone, designated NA-2-2, hybridized to both these probes. Figure 3A shows the nucleotide sequence of the cDNA insert from NA-2-2. The cDNA sequence at the 5' end of the NA-2-2 insert is identical to the sequence of the cDNA fragment synthesized by primer extension (Fig. 2C), and includes an ATG that could act as an initiation codon. In addition, the sequence of the NA-2-2 insert matches exactly the sequence of the NA-2-1 insert from nucleotide 19 of the NA-2-2 cDNA (Fig. 3A). The NA-2-2 sequence contains an open reading frame of 642 base pairs (bp) which encodes a protein with a predicted M_r of 24,847, including a putative signal sequence of 22 amino acids. Figure 3A shows the sequence regions that have been confirmed by sequencing of peptides generated from the S_2 -glycoprotein by either CNBr cleavage or digestion with *Staphylococcus aureus* protease. Of the predicted 192 amino-acid residues of the mature S_2 -glycoprotein, 61 have been confirmed by amino-acid sequencing of peptides. We therefore conclude that the sequence shown in Fig. 3A encodes the mature S_2 protein and a signal sequence that directs the transfer of the S_2 -glycoprotein from the transmitting tract cells to the extracellular space. This putative signal sequence has features that are typical of eukaryotic signal sequences²³—a relatively hydrophilic sequence of five amino acids at the C-terminal end, a relatively hydrophobic sequence (–6 to –14) and a more hydrophilic sequence at the N-terminal end (see Fig. 3B). The amino acids at the C-terminal end include glycine at the cleavage site, an aromatic residue at –2 and a helix breaker, proline, at –4, all of which are common features of the C-terminal region of signal sequences. In the mature protein sequence there is

relatively hydrophobic sequence at the N-terminus (Fig. 3B) which is homologous with that of *S*-allele-associated style components from *L. peruvianum*²⁰ (see Fig. 6). There are also three predicted glycosylation sites (residues 27, 37 and 138), based on the consensus glycosylation sequence (Asn-X-Ser/Thr)²⁴, in the hydrophilic region of the protein. As six moles of *n*-acetylglucosamine were detected per mole of isolated S_2 -glycoprotein (unpublished observations), it is possible that each of these glycosylation sites carries an *N*-linked carbohydrate chain. The fact that enzymatic deglycosylation with Endo F (Fig. 2A) yielded three products separable by SDS-polyacrylamide gel electrophoresis, suggests that these sites are in fact *N*-glycosylated. As well as these potential *N*-linked carbohydrate chains, there may also be *O*-linked chains. As no hydroxyproline was detected in the amino-acid analysis (unpublished observations), there is no opportunity for Hyp-glycosylation. As there are nine cysteine residues in the mature protein, there is considerable potential for internal sulphhydryl linkages as well as at least one free sulphhydryl group which might be functionally significant.

Apart from differences at the 5' end, clones NA-2-1 and NA-2-2 also differed in the lengths of their 3'-untranslated sequences. The two clones were identical as far as nucleotide 754, the polyadenylation site in clone NA-2-2; from this point the cDNA insert from clone NA-2-1 had an additional 50 nucleotides of untranslated mRNA and a poly(A) tail of 18 residues (see Fig. 3A). This difference at the 3' end suggests that there are alternative polyadenylation sites in S_2 RNA transcripts. The putative signals for the two sites are indicated in Fig. 3A; neither has the sequence AATAAA which is typically found in animal messages^{25,26}, but the polyadenylation signals in plant genes apparently show much greater sequence variation than is found in animal genes^{27,28}.

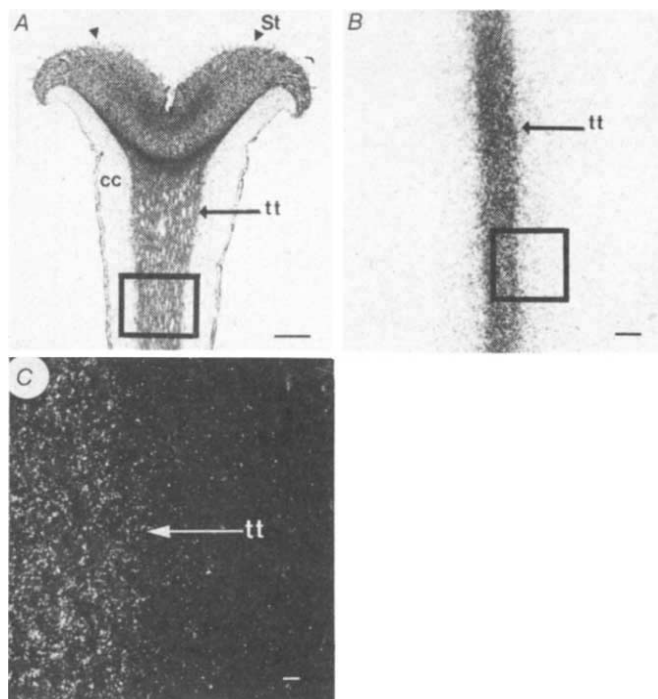
Hybridization analyses

The S_2 cDNA was used as a probe in Northern blot hybridization experiments with mRNAs prepared from mature styles of *N. alata* genotypes S_1S_3 , S_2S_3 , S_2S_2 and S_3S_3 , as well as from mature styles of *L. peruvianum* genotype S_1S_3 and green-bud styles and ovaries of *N. alata* genotype S_2S_3 (Fig. 4). In all mature styles bearing the S_2 allele, a predominant mRNA species of ~940 nucleotides hybridized to the cDNA probe. A mRNA of identical size was detected in preparations of poly(A)⁺ mRNA isolated from S_3S_3 and S_1S_3 styles of *N. alata*, the extent of hybridization being $\leq 1\%$ the level obtained in styles bearing the S_2 allele. This result indicates that the transcripts of the different *S* alleles are homologous. The same probe did not hybridize to green-bud mRNA; however, a weak signal corresponding to 940 nucleotides was detected in S_2S_3 ovary mRNA.

Self-incompatibility is not expressed in green buds, nor is the S_2 -glycoprotein detected in extracts of green buds; it is therefore not surprising that the S_2 cDNA probe did not hybridize with green-bud mRNA. The hybridization with ovary mRNA was unexpected as the S_2 -glycoprotein is not detected in extracts of ovary; it is possible that low levels of the glycoprotein are produced in the ovary to ensure arrest of any self-pollen tubes that manage to evade the self-incompatibility mechanisms expressed in the style. The RNA blot analysis also revealed that styles from *L. peruvianum* (genotype S_1S_3) synthesized readily detectable levels of a 2.5-kilobase (kb) mRNA that hybridized with the NA-2-1 cDNA insert. This finding is consistent with the extensive homology observed at the N-terminal sequences of the S_2 -glycoprotein from *N. alata* and both the S_1 and S_3 proteins from *L. peruvianum* (Fig. 6). The S_1 and S_3 proteins from *L. peruvianum* each have an estimated M_r of 28,000 and the RNA blot analysis indicated that the mRNA transcripts encoding these proteins are also identical in size. The predicted proportion of noncoding mRNA in the *L. peruvianum* S_1 and S_3 transcripts is significantly greater than for the S_2 transcript from *N. alata*.

Fig. 5 Hybridization histochemistry of probe for cDNA encoding the 32K style glycoprotein from *N. alata*, genotype S_2S_3 . **A**, Light micrograph of longitudinal section through a mature style (genotype S_2S_3) embedded in JB-4 resin (Polysciences) and stained with Richardson's blue. The central core of transmitting tissue (tt) is surrounded by the cortex (cc). The stigma surface (St) is indicated. Scale bar, 250 μm . **B**, Autoradiograph of a frozen longitudinal section through the style after hybridization with ^{32}P -labelled S_2 cDNA (obtained by direct exposure to X-ray film). Scale bar, 100 μm . The area shown corresponds to the boxed area in **A**. **C**, Darkfield photograph of a longitudinal section through the style shown in **B** after hybridization (obtained by liquid emulsion autoradiography). Scale bar, 10 μm . The area shown corresponds to the boxed section **B**.

Methods. Mature styles of *N. alata* (genotypes S_2S_3 and S_1S_3) were placed in a mould with OCT Tissue Tec and frozen in Freon 22 cooled in liquid nitrogen. Longitudinal sections (6–8 μm thick) were cut at -20°C using a Frigocut microtome (Reichert Jung) and attached to glass slides precoated with 1% gelatin using 0.25% formaldehyde as a crosslinking agent. The sections were fixed at 4°C for 5–10 min in 2% glutaraldehyde in 33% ethylene glycol and 0.1 M sodium phosphate buffer. Prehybridization and hybridization conditions were as described by Rall *et al.*⁴⁶. The ^{32}P -labelled S_2 cDNA (5×10^8 c.p.m. μg^{-1}) was hybridized to the sections at 40°C for 60 h; direct exposure time was 2 days using Cronex MRF 31 Medical Recording Film (Du Pont); exposure with liquid emulsion (Emulsion K5, Ilford Nuclear Research) was 7 days.



Hybridization to style sections

The cDNA encoding the S_2 -glycoprotein was also used in hybridization histochemical experiments using longitudinal sections of mature S_2S_3 and S_1S_3 styles. Figure 5 shows that there was specific binding of the cDNA probe to the cells of the S_2S_3 transmitting tissue, with no binding to the cortical or epidermal cells. The transmitting tract is the tissue through which the pollen tubes grow on their way to the ovary and it is in this tissue that we would expect the S gene to be expressed. The cDNA probe also bound to the transmitting tissue of S_1S_3 styles; although identical conditions were used, binding was very faint compared with that obtained using mature S_2S_3 styles. This result is consistent with the weak hybridization detected in the Northern blot experiments. In control experiments there was essentially no hybridization.

Homology between *N. alata* and *L. peruvianum*

N-terminal sequence analysis of the proteins associated with *L. peruvianum* self-incompatibility alleles S_1 and S_3 revealed striking homologies to the N-terminus of the *N. alata* S_2 -glycoprotein²⁰. The proteins associated with the S_1 and S_3 alleles of *L. peruvianum* were identified by electrophoretic analysis of style extracts and were purified by electroelution and sequenced as described by Mau *et al.*²⁰. Figure 6 shows the extent of homology between the *N. alata* S_2 -glycoprotein and the two S -associated proteins isolated from *L. peruvianum*. The homology at the N-terminal ends of the S -allele-associated style components of *N. alata* and *L. peruvianum* is consistent with the hypothesis that the S gene has been conserved in evolution to prevent inbreeding of plant species¹.

Discussion

We have cloned the cDNA encoding a glycoprotein from the female sexual tissue of *N. alata* which segregates with the S_2 allele of the gene controlling self-incompatibility. This cDNA hybridizes with mRNAs from styles of other S genotypes of *N. alata* and also with mRNAs from styles of *L. peruvianum*, another species in the family Solanaceae which has a gametophytic self-incompatibility system. Specific homology in the N-terminal sequences of S -genotype-related style proteins

for these two species has also been demonstrated.

Taken together, the segregation of the S_2 -glycoprotein with function of the S_2 allele, the *in vitro* activity of the isolated S_2 -glycoprotein as a pollen tube inhibitor²⁹, and the homology at the N-terminus of S -genotype-associated style components from *N. alata* and *L. peruvianum*, provide strong evidence that the cDNA described here encodes the product of the S_2 allele. However, conclusive evidence can only come from transformation of the S genotype by gene transfer.

The S gene is considered to have a tripartite structure—two conserved portions encoding pollen activity and style activity and a variable part encoding the allelic specificity¹. The conservation of the hydrophobic N-terminal sequences of the S -allele-associated components from *N. alata* and *L. peruvianum*²⁰ is compatible with this model; these sequences may represent a fragment of the conserved portion encoding style activity. The S_2 cDNA can now be used as a probe to obtain cDNA clones for other S -alleles from *N. alata* as well as other related species, an approach which should elucidate the expression of different allelic specificities. Although it is likely that the allelic differences are expressed as differences in the protein component of the style glycoprotein, the possibility that the allelic differences reside in the glycosyl moiety cannot be excluded. In this case, the allelic differences would be expressed primarily as glycosyl transferases of different specificities, and differences would then be detected in the structures of the glycosyl moieties of the products of different S alleles. Analysis of the carbohydrate moiety of the S -glycoprotein is presently limited by difficulties in obtaining sufficient material.

Comparison of the *N. alata* S_2 cDNA sequence with nucleotide sequences encoding S -genotype-associated proteins from female tissues of other plant species may provide some insight into the evolution of self-incompatibility. Recently, a cDNA clone encoding part (~30%) of an S -locus-specific glycoprotein from *Brassica oleracea*, a species having a sporophytic self-incompatibility system¹⁵, has been described; there is no apparent homology between this partial sequence and that of the *N. alata* S_2 cDNA. However, the possibility of homology with the other undescribed 70% of the molecule remains. Once a full-length clone for the *B. oleracea* protein is available, comparison of the two sequences should give a greater under-

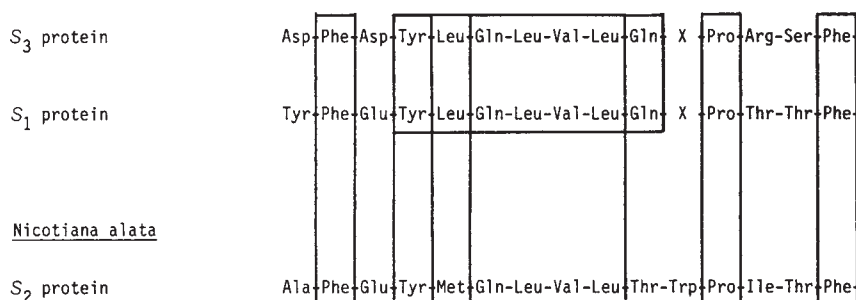
Lycopersicon peruvianum

Fig. 6 A comparison of the amino-terminal sequences of *L. peruvianum* S_1 - and S_3 -associated proteins with *N. alata* S_2 protein. Identical amino acids are boxed.

standing of the evolutionary relationship between the sporophytic and gametophytic systems.

A fundamental question in understanding how self-incompatibility is expressed is whether the pollen tube and style S -gene products are identical. To date, we have not been able to identify a protein from pollen which corresponds to the S genotype. There are two reasons for expecting the products of a particular S allele to differ in pollen and style: first, it is difficult to envisage how two identical molecules could interact to cause inhibition of pollen tube growth. It is possible that the S -gene products from the style and pollen tube, although generated from the same allele, are not identical. Non-identical products could be generated by different splicing of the same transcriptional unit to produce different mRNAs, or there may be different post-translational modifications such as glycosylation²⁹. In this way, differences leading to complementary sites on the two S -gene products might be generated. The second reason for expecting differences in the structures of the two S -gene products relates to their secretion. It is assumed that the cells of the style transmitting tract secrete the S -gene product into the extracellular matrix; the presence of the signal sequence predicted by the S_2 cDNA sequence is compatible with secretion of the product. The pollen tubes grow through the style in the extracellular matrix and must in some way recognize the S -gene product. It is difficult to envisage how the recognition which results in pollen tube arrest could occur if the pollen S -gene product, as

well as the style S -gene product, is secreted. It is more likely that the pollen S -gene product is either membrane-associated or wall-associated, so that the interaction with the style S -gene product could signal cytoplasmic events leading to arrest of pollen-tube growth.

The cloning of the cDNA encoding the S_2 -glycoprotein represents the initial step in resolving these questions, which have broader implications in the field of cell-cell recognition in plants. It is likely that an understanding of how allelic differences are expressed by plant genes and how products of these alleles interact to initiate a response involving changes in cell-wall metabolism, will lead to principles which may apply to more complex systems involving plant cell-cell interactions, for example, in host-pathogen and host-parasite interactions.

We thank Dr K. K. Pandey (Grasslands, Palmerston North, New Zealand), for providing *N. alata* seed of genotypes S_2S_3 and S_1S_3 , Mr Darron Broom for management of the glasshouse and Ms Susan Mau for collection of plant material. We also thank Lucy Duncan for assistance in preparation of the synthetic oligonucleotide probes; Marie John for assistance with amino-acid sequencing; Dr Tony Bacic for the monosaccharide analyses; Marg Strode for preparation of the manuscript; and Tony Kyne for making the MELBDBSYS suite of computer programs available to us. The work was supported in part by funds from Agrigenetics Research Associates.

Received 21 October 1985; accepted 12 March 1986.

- de Nettancourt, D. *Incompatibility in Angiosperms* (Springer, Berlin, 1977).
- Heslop-Harrison, J. *Proc. R. Soc. B* **202**, 73-92 (1978).
- Lewis, D. *N.Z. J. Bot.* **17**, 637-644 (1979).
- Pandey, K. K. *N.Z. J. Bot.* **17**, 645-663 (1979).
- Harris, P. J., Anderson, M. A., Bacic, A. & Clarke, A. E. *Oxford Surveys of Plant Molecular and Cell Biology* Vol. 1 (ed. Milfin, B. J.) 161-203 (Oxford University Press, 1984).
- Knox, R. B. in *Encyclopaedia of Plant Physiology, New Series* Vol. 17 (eds Linskens, H. F. & Heslop-Harrison, J.) 508-608 (Springer, Berlin, 1984).
- Larsen, K. *Heredity* **85**, 227-248 (1977).
- Pandey, K. K. *Heredity* **22**, 255-283 (1967).
- Lawrence, M. J. *Proc. R. Soc. B* **188**, 275-285 (1975).
- Lawrence, M. J., Afzal, M. & Kenrick, J. *Heredity* **40**, 239-253 (1978).
- Linskens, H. F. *Proc. R. Soc. B* **188**, 299-311 (1975).
- Nishio, T. & Hinata, K. *Jap. J. Genet.* **54**, 307-311 (1979); *Genetics* **100**, 641-647 (1982).
- Ferrari, T. E., Bruns, D. & Wallace, D. H. *Pl. Physiol.* **67**, 270-277 (1981).
- Mau, S.-L., Raff, J. & Clarke, A. E. *Planta* **156**, 505-516 (1982).
- Nasrallah, J. B., Kao, T.-H., Goldberg, M. L. & Nasrallah, M. E. *Nature* **318**, 263-267 (1985).
- Lewis, D. *Heredity* **3**, 339-345 (1949); *Proc. R. Soc. B* **151**, 468-477 (1960); in *Genetics Today* Vol. 3 (ed. Geerts, S. J.) 656-663 (Pergamon, Oxford, 1965).
- Nasrallah, M. E., Barber, J. T. & Wallace, D. H. *Heredity* **24**, 23-27 (1969).
- Heslop-Harrison, J. A. *Rev. Pl. Physiol.* **26**, 403-425 (1975); *Proc. R. Soc. B* **218**, 371-395 (1983).
- Burnet, F. M. *Nature* **232**, 230 (1971).
- Mau, S.-L. *et al. Planta* (submitted).
- Bredemeijer, G. M. M. & Blass, J. *Theor. appl. Genet.* **59**, 185-190 (1981).
- Okayama, H. & Berg, P. *Molec. cell. Biol.* **2**, 161-170 (1982).
- von Heijne, G. *Eur. J. Biochem.* **133**, 17-21 (1983); *J. molec. Biol.* **184**, 99-105 (1985).

- Wagh, P. V. & Bahl, O. P. *CRC Crit. Rev. Biochem.* **10**, 307-377 (1981).
- Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211-214 (1976).
- Beinoist, C., O'Hare, K., Beathnach, R. & Chambon, P. *Nucleic Acids Res.* **8**, 127-140 (1980).
- Lycett, G. W., Delauney, A. J. & Croy, R. R. D. *FEBS Lett.* **153**, 43-46 (1983).
- Messing, J. *et al.* in *Genetic Engineering of Plants (Basic Life Sci. Ser.)* (eds Kosuge, T., Meredith, C. P. & Hollaender, A.) 211-227 (Plenum, New York, 1983).
- Clarke, A. E., Anderson, M. A., Bacic, T., Harris, P. J. & Mau, S.-L. *J. Cell Sci. Suppl.* **2**, 261-285 (1985).
- Bradford, M. M. *Analyt. Biochem.* **72**, 248-253 (1976).
- Edge, A. S. B., Faltynek, C. R., Hof, L., Reichert, L. E. & Weber, P. *Analyt. Biochem.* **118**, 131-137 (1981).
- Malmberg, R., Messing, J. & Sussex, I. in *Molecular Biology of Plants: A Laboratory Course Manual* (Cold Spring Harbor Laboratory, New York, 1985).
- Beaucage, S. L. & Caruthers, M. H. *Tetrahedron Lett.* **22**, 1859-1862 (1981).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning: A Practical Approach* (ed. Glover, D.) (IRL, Oxford, 1984).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).
- Messing, J. & Vieira, J. *Gene* **19**, 269-276 (1982).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161-178 (1980).
- Staden, R. *Nucleic Acids Res.* **4**, 4037-4051 (1977); *Nucleic Acids Res.* **5**, 1013-1015 (1978).
- Grego, B. *et al. Eur. J. Biochem.* **148**, 485-491 (1985).
- Simpson, R. J. & Nice, E. C. *Biochem. Int.* **8**, 787-791 (1984).
- Nice, E., Grego, B. & Simpson, R. J. *Biochem. Int.* **11**, 187-195 (1985).
- Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Rall, L. B. *et al. Nature* **313**, 228-231 (1985).

'Missing image' in gravitational lens systems?

D. Narasimha*, K. Subramanian† & S. M. Chitre‡

* Department of Physics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

† Astronomy Centre, University of Sussex, Brighton BN1 9QH, UK

‡ Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

Gravitational lens models involving extended mass distributions generally predict an odd number of images¹, with one of the images close to the centre of the principal lens galaxy. In all the lens systems observed up to now, only an even number of images have been unambiguously detected. A number of suggestions²⁻⁴ have been advanced to explain the absence of the 'odd' image, but none of these is sufficiently compelling. We suggest here that the presence of a compact nucleus with mass in the range 5×10^9 to 5×10^{10} solar masses ($M_\odot$), contained within a region of size ≤ 50 pc at the centre of a lens galaxy, would dim the odd image significantly without affecting the rest of the image configuration.

We argue that a massive compact nucleus with even a small fraction ($\leq 2\%$) of the total galaxy mass contained within ~ 50 pc would have negligible influence on the properties of the images located outside the core of the galaxy. But the proximity of the odd image to the nucleus allows it to be pulled closer to the galactic centre, at the same time becoming significantly dimmer. We demonstrate below that the effect becomes pronounced for a nuclear mass exceeding $\sim 5 \times 10^9 M_\odot$ contained within ~ 50 pc. It turns out that the exact size of the nucleus is not particularly important, although the intensity of the odd image increases as the nucleus becomes more extended.

Consider a spherically symmetric density distribution of matter for the core of the underlying lens galaxy. The scattering function⁵ representing the bending of light rays for an impact parameter small compared with the radius of the galactic core is given by

$$I_G(Z_0) = \pi \sigma_G Z_0^* \quad (1)$$

Here, in the usual notation, σ_G is the central surface density of the galaxy and Z_0 is the complex number giving the image position in the deflector plane. Suppose, in addition we consider the effect of a nucleus of mass M_N and radius R_N , with a constant surface density. Then the scattering function of the nucleus can be written as

$$\begin{aligned} I_N(Z_0) &= \frac{M_N}{R_N^2} Z_0^*, \quad |Z_0| < R_N \\ &= M_N / Z_0, \quad |Z_0| > R_N \end{aligned} \quad (2)$$

With the help of the above scattering functions we can relate the image position Z_0 to the source position Z , through the equation⁵

$$\begin{aligned} Z &= Z_0 - \frac{4GD}{c^2} I_G^* - \frac{4GD}{c^2} I_N^* \\ &= Z_0(1 - \kappa_G - \kappa_N), \quad |Z| < R_N \end{aligned} \quad (3)$$

where

$$\kappa_G = \frac{4GD}{c^2} \pi \sigma_G, \quad \kappa_N = \frac{4GD}{c^2} \pi \left(\frac{M_N}{\pi R_N^2} \right) = \frac{R_0^2}{R_N^2} \quad (4)$$

$R_0 = [(4GD/c^2)M_N]^{1/2}$ is the radius of influence of the nucleus and $D = DD \times DDS/DS$, where DD , DDS and DS are the distances between observer and deflector, deflector and source,

Table 1 Image properties in the presence of the galactic nucleus

Lens system	$M_N (M_\odot)$	$M_G (M_\odot)$	Intensity ratio		Ref.
			Without nucleus	With nucleus	
0957+561	1.4×10^{10}	4.8×10^{11}	2.5×10^{-2}	1.0×10^{-6}	6
2345+007	1.3×10^{10}	1.4×10^{12}	1.8×10^{-2}	8.0×10^{-5}	7
2016+112	2.5×10^{10}	3.4×10^{12}	1.5×10^{-2}	4.0×10^{-6}	8

M_N , mass of galactic nucleus; M_G , mass of parent lens galaxy (both in units of solar mass, $M_\odot$). The intensity ratio is that between the odd (third) image and the external bright image.

and observer and source, respectively. G is the gravitational constant and c is the speed of light in vacuum.

In the absence of the nucleus the image would have been located at a position Z_0^1 given by

$$Z = Z_0^1 (1 - \kappa_G) \quad (5)$$

Let $Z = \gamma e^{i\theta}$, $Z_0 = \gamma_0 e^{i\phi}$ and $Z_0^1 = \gamma_0^1 e^{i\psi}$. In the context of the present problem, as will become clear later, both κ_G and κ_N exceed unity. Then, without the presence of the nucleus, the 'odd image' is located at

$$\gamma_0^1 = \frac{\gamma}{(\kappa_G - 1)} \quad (6)$$

The introduction of the nucleus shifts the image to the position

$$\gamma_0 = \frac{\gamma}{(\kappa_G + \kappa_N - 1)} = \frac{(\kappa_G - 1) \gamma_0^1}{(\kappa_G + \kappa_N - 1)} \quad (7)$$

Note that for values of κ_G exceeding unity and for the location, in the absence of the compact nucleus, of the odd image γ_0^1 within a critical radius, $\approx \kappa_G^{1/2} R_0$, the nucleus does not create additional sub-images of the odd image (ref. 3, section 2(e), (f)). The amplification A^1 of the image in the absence of the nucleus is obtained using equation (6):

$$A^1 = \frac{\gamma_0^1 d\gamma_0^1}{\gamma d\gamma} = \frac{1}{(\kappa_G - 1)^2} \quad (8)$$

whereas after the introduction of the nucleus the amplification A follows from equation (7):

$$A = \frac{\gamma_0 d\gamma_0}{\gamma d\gamma} = \frac{1}{(\kappa_G + \kappa_N - 1)^2} \quad (9)$$

Suppose the parent galaxy has a mass $M \approx 5 \times 10^{11} M_\odot$, within a core-radius ≈ 3 kpc. Assuming $D \approx 600$ Mpc, we get $\kappa_G \approx 6.5$. If we now consider a nucleus with mass $M_N \approx 10^{10} M_\odot$ and radius $R_N \approx 50$ pc, then we have $\kappa_N \approx 470$. Clearly κ_N is very large compared with κ_G , so that from equations (4) and (7) we have

$$\gamma_0 \approx \frac{\kappa_G}{\kappa_N} \gamma_0^1 = \kappa_G \left(\frac{R_N}{R_0} \right) \left(\frac{\gamma_0^1}{R_0} \right) R_N \quad (10)$$

Thus for $\gamma_0^1 \approx \kappa_G^{1/2} R_0$ and for the foregoing choice of parameters, the image in the presence of the nucleus is located at

$$\gamma_0 \approx \kappa_G^{3/2} \left(\frac{R_N}{R_0} \right) R_N \approx \frac{3}{4} R_N.$$

The amplifications before and after the introduction of the nucleus are given by equations (8) and (9):

$$A^1 = 3 \times 10^{-2}; \quad A = 4 \times 10^{-6} \quad (11)$$

It is evident from this idealized example that after the introduction of the nucleus the odd image is depressed in intensity by a factor of $\sim 10^4$, or equivalently dimmed by ~ 10 mag, in addition to being brought closer to the centre of the galaxy (and the nucleus) by a factor of ~ 80 .

We now incorporate the effects of the compact nucleus on the image properties by adopting the prescription outlined for

realistic lens models⁶. In fact, we merely add the gravitational effect of the compact nucleus to the lens models which we have previously constructed⁶⁻⁸, without altering any of the parameters of those models.

Table 1 summarizes the numerical results for three multiple quasar systems: 0957+561, 2345+007 and 2016+112. In these models, without the inclusion of the effects of the nucleus the intensity of the odd (third) image is a few percent of that of the external bright image. But in the presence of the nucleus the intensity of the third image is enormously diminished, and the intensity ratio is lowered by a factor anywhere from 10^2 to 10^4 ; that is, the odd third image is dimmed by $\sim 6-9$ mag. A remarkable feature is that the properties of the main images, such as their positions and intensities, are left practically unaltered by the presence of the nucleus, while the third image is pulled significantly closer to the centre of the nucleus. The masses of the parent lens galaxy required to obtain the observed configuration are in the range 5×10^{11} – $3.4 \times 10^{12} M_\odot$, with the mass of the nucleus $\approx 1.3 \times 10^{10}$ – $2.5 \times 10^{10} M_\odot$ contained within ~ 50 pc. Further, the intensity of the third image would increase if the nucleus is assumed to be more extended or the mass of the nucleus lower than the values given in Table 1.

We note that the present model does not necessarily involve any singularities (for example, black holes or strings) in the density distribution of matter, although such singularities can also play a similar role in explaining the absence of the odd image. We consider that it is better to try to account for most of the observed features of the gravitational lens system, including the missing odd image, in the framework of a conventional, smooth density distribution, before invoking the presence of singularities. Indeed, Young *et al.*⁹ have inferred from photometric and spectroscopic observations that the nucleus of the giant elliptical galaxy M87 contains a compact mass $\approx 5 \times 10^9 M_\odot$ within 100 pc (see also Sargent *et al.*¹⁴).

The presence of a black hole in the central region of a lens galaxy could cause the complete disappearance of the third image, due to the action of the black hole as a 'mini-lens'. In contrast, a massive compact nucleus will merely dim the image by a substantial amount⁵. The dimmed image could be detected if the source size (for example, the extended narrow-emission-line region in the quasar) is comparable to the radius of influence of the nucleus, for, in such a situation, some fraction of the large emission-line region is unlikely to be significantly affected by the lens action of the massive nucleus.

Will the existence of clumpiness in the lens affect the intensity of the odd image? For parameters appropriate to the present problem and for clumps in the form of solar-mass stars, neither the overall bending nor the intensity amplification of the large radio and emission-line regions are sensibly influenced. But for the compact optical continuum region we cannot neglect the effects of minilensing. However, in such a situation arguments^{10,11} based on flux conservation and the large optical depth to mini-lens action¹² enable us to assign a probability distribution¹³, $P(A)$, to the intensity amplification of the odd image which peaks at $A \approx 4 \times 10^{-6}$. As a result the clumpiness in the lens galaxy or the nucleus is not expected to affect our conclusions about the dimness of the odd image.

A testable prediction of the present suggestion is the observation of an image of the quasar optical continuum region within ~ 0.2 arc s from the centre of the lens galaxy, with visual magnitude fainter by nearly 6 mag than the external bright image. Equally, if the source size is large compared with R_0 in some frequency, such as the narrow-emission-line region, then in the present model the image could still be observed near the centre of the lens galaxy. On the other hand, the detection of a bright odd image in the optical continuum will invalidate the hypothesized presence of a compact massive nucleus in the core of the lens galaxy and would imply that the nucleus is either less massive than $5 \times 10^9 M_\odot$ or has a linear size larger than 50 pc.

The numerical computations were carried out on the CYBER

175 and CDC 205 computers at the University of Calgary. D.N. was supported by a Canadian NSERC grant to S. R. Sreenivasan. K.S. was supported by SERC.

Received 2 January; accepted 25 February 1986.

1. Burke, W. L. *Astrophys. J.* **244**, L1-LX (1981).
2. Chang, K. & Refsdal, S. *Astr. Astrophys.* **132**, 168-178 (1984).
3. Subramanian, K., Chitre, S. M. & Narasimha, D. *Astrophys. J.* **289**, 37-51 (1985).
4. Narayan, R., Blandford, R. & Nityananda, R. *Nature* **310**, 112 (1984).
5. Bourassa, R. R. & Kantowski, R. *Astrophys. J.* **195**, 13-21 (1975).
6. Narasimha, D., Subramanian, K. & Chitre, S. M. *Mon. Not. R. astr. Soc.* **210**, 79-83 (1984).
7. Subramanian, K. & Chitre, S. M. *Astrophys. J.* **276**, 440-448 (1984).
8. Narasimha, D., Subramanian, K. & Chitre, S. M. preprint, Tata Institute, Bombay (1985).
9. Young, P. J., Westphal, J. A., Kristian, J., Wilson, C. P. & Landauer, F. P. *Astrophys. J.* **221**, 721-730 (1978).
10. Weinberg, S. *Gravitation and Cosmology* (Freeman, San Francisco, 1972).
11. Cowling, S. A. preprint, Institute of Astronomy, Cambridge (1985).
12. Vietri, M. & Ostriker, J. P. *Astrophys. J.* **267**, 488-510 (1983).
13. Paczynski, B. preprint, Princeton University Observatory (1985).
14. Sargent, W. L. W., Young, P. J., Boksenberg, A., Shortridge, K., Lynds, C. R. & Hartwick, F. D. A., *Astrophys. J.* **221**, 731-744 (1978).

Resolution of Crab nebula filaments into chains of quasistellar knots

Sidney van den Bergh* and Christopher J. Pritchett†

* Dominion Astrophysical Observatory,
Herzberg Institute of Astrophysics, 5071 W. Saanich Road, Victoria,
British Columbia, Canada V8X 4M6

† Department of Physics, University of Victoria, PO Box 1700,
Victoria, British Columbia, Canada V8W 2Y2

We have clearly resolved a number of the $H\alpha$ emitting filaments in the Crab nebula into chains of knots with diameters ≤ 1.0 arc s. This resolution was possible because sub-arcsecond seeing is common on Mauna Kea, Hawaii (altitude 4,204 m) but quite rare at the sites of other large telescopes situated at lower altitudes. Individual filaments resemble beads on a string which shows that these objects are rope-like structures. In the brightest filament close to the Crab pulsar the quasistellar knots are slightly displaced from the spine of the low-density filament on which they are superposed. This segregation might be understood if the dense knots have been accelerated outwards by a lesser amount than the low-density filament in which they were originally embedded. The present results show that a substantial fraction of the line emission of the Crab nebula originates in the small and very dense quasistellar knots that lie along the spines of filaments. This will affect estimates of the mass and composition of the Crab nebula, and observations of the motions of quasistellar knots should enable one to study the expansion (and eventually perhaps even the acceleration) of the Crab nebula with much greater precision than was possible previously.

On 12 October 1985 (UT) we obtained frames of two fields in the Crab nebula with the 320×512 pixel RCA charge-coupled device (CCD) detector at the prime focus of the Canada-France-Hawaii Telescope. These 600-s exposures were made in 0.8-arc-s seeing through a ~ 100 -Å-wide filter centred on $H\alpha$. The scale of the resulting images was 0.41 arc s per pixel. On these frames many of the $H\alpha$ emission filaments resemble necklaces (see Figs 1 and 2) composed of 'beads' of differing sizes. Some of these beads appear stellar whereas others are barely resolved with diameters ~ 1.0 arc s. Typically such knots are strung out along the spines of individual filaments at separations of a few seconds of arc. Bright and faint filaments appear to have the same basic structure. However, in bright filaments individual knots tend to be more luminous (and more closely spaced) than they are in fainter filaments. The appearance of the $H\alpha$ emission filaments, and the fact that most knots are located in filaments, suggests that they are (in three dimensions) rope-like structures rather than sheets seen edge-on. One might

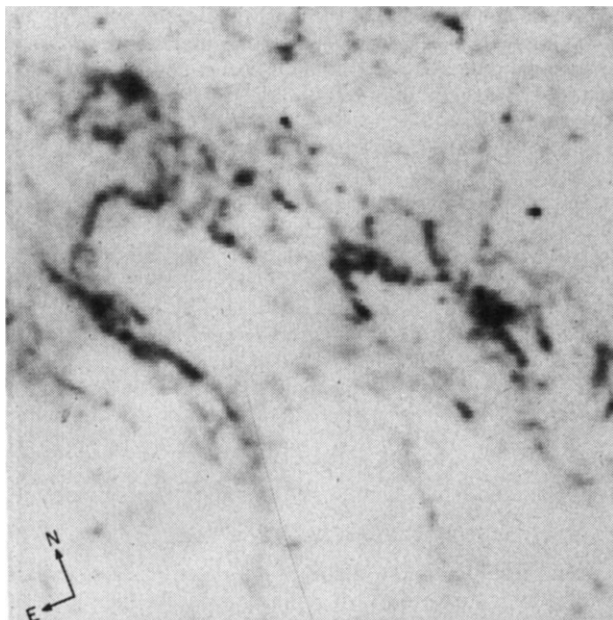


Fig. 1 H α filament located north of centre of the Crab nebula. The filament is seen to be resolved into a string of quasistellar knots with diameters ≤ 1.0 arc s. The figure is 60 arc s on a side. Pixel size is 0.41 arc s.

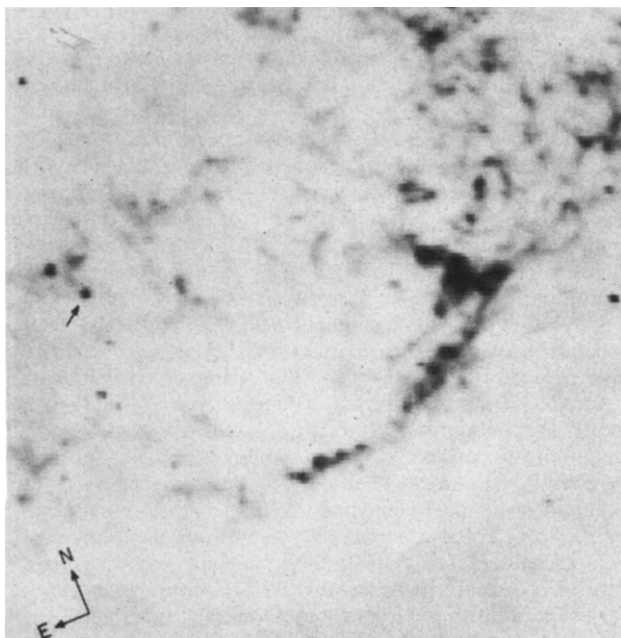


Fig. 2 Filament located west of the Crab pulsar, which is marked by an arrow. This frame was taken in 0.8-arc-s seeing. The figure is 60 arc s on a side.

speculate that individual knots are due to instabilities within these filaments. The distribution of these knots and the pressures needed to form and confine them should provide information about the otherwise unprobed small-scale distribution and properties of the hot thermal plasma and magnetic field between and around the filaments.

Our CCD frames cover two 2×3 -arc-min fields. One of these was centred on the Crab pulsar and the other is located slightly to the west of the base of the Crab jet¹. We estimate that the total number of H α emission knots in the Crab nebula is a few hundred. At an assumed distance^{2,3} of 2 kpc the ≤ 1.0 -arc-s diameters of individual knots correspond to linear dimensions ≤ 0.01 pc. The total emitting volume of all knots in the Crab

nebula is therefore only a few $\times 10^{-4}$ pc³. A substantial fraction of the total H α emission of the Crab nebula must originate within this very small volume.

Important questions that remain to be answered by future observations are: (1) Do the quasistellar H α knots also radiate in the light of elements other than hydrogen? (2) What is the range of physical conditions within these knots? (3) What are the lifetimes of quasistellar knots? (4) How do the spectra of knots differ from those of the filaments in which they are embedded?

The observations reported above were obtained with the Canada-France-Hawaii Telescope, which is operated by the NRC of Canada, CNRS of France and the University of Hawaii. We thank Peter Stetson and David Duncan for their assistance with the processing and reproduction of the digital images shown in the figures.

Received 6 January; accepted 5 March 1986.

1. van den Bergh, S. *Astrophys. J. Lett.* **160**, L27 (1970).
2. Woltjer, L. *Bull. astr. Inst. Neth.* **14**, 39-80 (1958).
3. Trimble, V. *Astr. J.* **73**, 535-547 (1968).

A flare-induced cascade model of γ -ray bursts

P. A. Sturrock

Center for Space Science and Astrophysics, Stanford University, Stanford, California 94305, USA

Although there is no agreement concerning the model and mechanisms responsible for γ -ray bursts, it is generally agreed that they are produced by neutron stars¹. With few exceptions², it is thought that an intense magnetic field plays a key role. The favoured radiation mechanism is optically thin synchrotron radiation³, but it is difficult to find any way of maintaining the electron energy since the radiation cooling time is extremely short. As a resolution of this difficulty, I propose that the basic energy-release mechanism is a flare in the magnetosphere of a neutron star. This involves reconnection that leads to an electric field parallel to the magnetic field. This accelerates an electron along the magnetic field, producing high-energy γ rays (by curvature radiation) that promptly annihilate in the magnetic field, resulting in an electron-positron cascade as in radio pulsars⁴. This model offers an explanation of the continuum spectrum, of the 511-keV annihilation line, and possibly of the optical radiation that seems to accompany some γ -ray bursts.

An electron of energy E_e (eV) moving in a magnetic field B (gauss) at a large pitch angle produces predominantly photons of energy E_ϕ (eV) where

$$E_\phi = 10^{-19.7} E_e^2 B \quad (1)$$

Hence if observed photons of energy 10^5 eV are in fact produced in a magnetic field of order 10^{12} gauss, the emitting electrons must have energy of the order of $10^{6.4}$ eV. Since the power S emitted by a single electron by this process is given by

$$S_1 = 10^{-26.1} E_e^2 B^2 \quad (2)$$

we see that the timescale characteristic of radiation cooling of the emitting electron is given by

$$\tau_R = 10^{14.4} E_e^{-1} B^{-2} \quad (3)$$

For the parameters of interest, τ_R is of the order of 10^{-16} s. The problem is to find a process that can energize electrons as rapidly as, or more rapidly than, they are being de-excited by synchrotron radiation.

Basically, the excitation must be achieved by an electric field. One possibility is that this electric field is intrinsic to the photon

flux itself, so that absorption of photons is going on at the same high rate as the emission of photons. This occurs only if the optical depth of the emitting region is of order unity or more, and this possibility is ruled out since observed spectra do not display a Wien peak. (On the other hand, as Hameury *et al.*⁵ have shown, excitation of electrons by a distinct 'primary' beam of photons that have *not* been produced by synchrotron radiation, *can* lead to optically thin synchrotron radiation. Their model is in some respects similar to the one to be presented herein.)

Another possibility is that the electric field required to excite an electron is provided by a close encounter with another charged particle—electron, positron, or ion. However, we find that the particles responsible for exciting the radiating electrons must have an enormous density, of order 10^{31} cm^{-3} or more. Clearly, the focus of the inquiry then changes to a study of radiation from that very dense plasma. Hence we have not solved a difficult problem, but merely replaced it by another problem more difficult still.

For the above reasons, it has been proposed^{3,6,7} that excitation of the electrons is maintained by a collective plasma process. This process must generate an intense electric field since, to maintain the electron energy against loss by synchrotron radiation for the parameters proposed, the electric field must have a value of $10^{9.5}$ e.s.u. or more. A further requirement is that the electric field must be transverse to the magnetic field and have a frequency comparable with the orbital frequency of the electrons in the magnetic field. For the parameters quoted, this frequency is of the order of $10^{17.8}$ Hz, corresponding to a photon energy of $10^{3.4}$ eV. If this were the case, we would expect the source to emit more energy in X rays than in γ rays, which is normally not the case.

It is clearly very difficult to propose an acceptable mechanism by which electrons may be energized as rapidly as they are being cooled by synchrotron radiation. As a resolution of this dilemma, I propose that energy lost by synchrotron radiation is not restored to the emitting electron.

The impulsive nature of γ -ray bursts and the 'light curves' of the γ -ray flux are reminiscent of solar flares⁸. It is not inconceivable that the energy released during a γ -ray burst has been stored as magnetic free energy in a neutron-star magnetosphere, so it is attractive to consider the consequences of a neutron-star flare. Although there are several models of solar flares, all current models involve magnetic-field reconnection by the tearing mode or a similar processes.⁹ Magnetic reconnection violates the 'frozen-flux' condition and therefore develops electric fields that have a non-zero component parallel to the magnetic field. We now examine the consequence of the development of such an electric field in a neutron-star magnetosphere. Since this same situation arises in most models of radio pulsars, we may draw upon pulsar theory in investigating the present problem.

In the presence of a strong electric field parallel to the magnetic field, electrons rapidly accelerate, but their acceleration is soon limited by curvature radiation⁴. The radiation-reaction limited energy, measured in eV, is given by

$$E_{\text{cRL}} = 10^{8.1} \mathcal{E}^{1/4} R_c^{1/2} \quad (5)$$

where $\mathcal{E}$ (e.s.u.) is the electric field strength and R_c (cm) is the radius of curvature. During magnetic reconnection, the strength of the electric field may approach $10^{-2}(v_A/c)B$, where v_A is the Alfvén speed¹⁰. In the present situation, the Alfvén speed is likely to be close to the speed of light so that $\mathcal{E}$ might well be of the order of 10^{10} e.s.u. Adopting $R_c \approx 10^6$, we then see that E_{cRL} will be of the order of $10^{13.6}$ eV. On using the formula

$$E_\phi = 10^{-22.2} E_e^3 R_c^{-1} \quad (6)$$

we see that these electrons will emit photons with energy of the order of $10^{12.6}$ eV by curvature radiation.

As in the case of radio pulsars, a photon of this energy would not escape from the magnetosphere but would rapidly annihilate

to produce an electron and a positron, each of which has approximately half the energy of the annihilated photon. As shown some time ago⁴, annihilation occurs approximately at a location where the photon 'sees' a transverse magnetic field $B_\perp$ that satisfies

$$E_\phi B_\perp = 10^{18.6} \quad (7)$$

The newly created electrons and positrons have non-zero pitch angles and so radiate by the synchrotron process. These photons also will have such high energy that they will be immediately annihilated by the magnetic field. This cascade ends when photons are produced that are of sufficiently low energy that they may escape. If the above process were to occur only in a small region where the magnetic field strength is constant, the emitted spectrum will be of limited extent. The lower limit of the spectrum, $E_{\phi,L}$, is given approximately by the gyrofrequency:

$$E_{\phi,L} \approx 10^{-7.9} B \quad (8)$$

The upper limit to the spectrum, $E_{\phi,U}$, is given approximately by

$$E_{\phi,U} = 10^{18.6} B^{-1} \quad (9)$$

Between these limits, the shape of the spectrum is essentially that of a single electron as it radiates away its energy by the synchrotron process. We find, from Equation (B.18) of ref. 4, that the spectrum of a single electron is given by

$$N_{\phi,1} = 10^{9.9} B^{-1/2} E_\phi^{-3/2} \quad (10)$$

where $N_{\phi,1} dE_\phi$ is the number of photons with energy in the range E_ϕ to $E_\phi + dE_\phi$. The γ -ray burst spectra compiled in refs 11 and 12 exhibit great variation. However, a few cases are found to approximate a power-law form over a substantial range of energy. In these cases, we find that the index is indeed close to -1.5 .

In most cases, it is likely that radiation will be produced by an extended volume of the magnetosphere in which the field strength varies over a wide range. The resultant spectrum will then differ from the above form, the actual spectrum depending on the magnetic-field geometry, and on the way that energy release is distributed.

One simple model is to assume that the flare occupies most of the volume of a dipole magnetic field and that the free energy density is proportional to B_0^2 , where B_0 is the potential field of the dipole. In this case, we find that the emergent gamma-ray spectrum is composed of three sections: below $E_{\phi,LM}$, the spectrum is flat; between $E_{\phi,LM}$ and $E_{\phi,UM}$, the slope is $-3/2$ as in equation (10); above $E_{\phi,UM}$, the slope is -3 . $E_{\phi,LM}$ and $E_{\phi,UM}$ are given by equations (8) and (9), with B replaced by B_M , the maximum value of the magnetic field strength in the magnetosphere. Some observed spectra fit this form reasonably well; the required values of B_M are found sometimes to be as high as 10^{13} gauss or more. It is possible that curvature radiation will make some contribution to the observed spectrum, and this may be the origin of the high-energy tail to the spectrum as measured in some cases by the γ -ray spectrometer on board the Solar Maximum Mission¹³.

If N_L is the total number of leptons (electrons and positrons) produced in the terminal stage of the cascade, the total spectrum of the fluence from the gamma-ray burst is given by

$$N_\phi = N_L N_{\phi,1} \quad (11)$$

However, N_L will also be the total number of photons emitted in the 511 keV line when the electrons and positrons annihilate. In some circumstances, the plasma will be so dense that annihilation occurs in the magnetosphere, and the line will be broad due both to particle velocities and the variation of gravitational potential energy. In other circumstances, the plasma will be sufficiently thin that most annihilation occurs at the surface of the star. In this case, the line should be fairly sharp. It is clear from the above formulas that comparison of the fluence in the annihilation line to the fluence in the continuum gives an esti-

mate of the mean strength of the magnetic field in the emitting region.

By using equation (11) to estimate the total number of electrons and positrons involved in a gamma-ray burst, and estimating the lifetime of the electrons either from *in situ* annihilation or flow to the surface of the star, we may estimate the electron-positron number density. This may vary over a wide range, but a typical value would be 10^{20} cm^{-3} . The fact that synchrotron radiation is optically thin implies that $\sigma_T n c < 1$, where σ_T is the Thomson cross-section and b is the instantaneous thickness of the radiating region of the magnetosphere. One is thus led to infer that b is less than about 10^4 cm —perhaps much less—but this is not unreasonable if energy release occurs in the neighbourhood of a thin (possibly moving and evolving) current sheet.

An electric field acting on an electron-positron plasma will tend to develop a two-stream flow. In solar physics, an electron stream moving at high speed through the corona develops plasma oscillations that then radiate at radio frequencies to produce a 'type III radio burst'¹⁴. In the present situation, the plasma frequency will be in the infrared, optical, or possibly (as Liang¹⁵ has already suggested) in the ultraviolet part of the spectrum. Since the radiation is produced by a collective process, the brightness temperature can be much higher than the plasma temperature. Hence the development of plasma oscillations and radiation from these oscillations could be responsible for the optical radiation that is believed to accompany some γ -ray bursts^{16,17}.

Many questions remain, the most pressing of which is the origin of the free magnetic energy released during the flare. We know that all large solar flares, and possibly all small ones also, are due to the interplay of two flux systems. It is believed that in some flares 'new' flux slowly intrudes into an 'old' flux system, followed at some later stage by reconnection¹⁸. On the other hand, it appears that in most large flares a magneto-hydrodynamic (MHD) process leads to the rapid injection of new flux into the old flux, and that 'driven' reconnection occurs during this process¹⁹. In the case of a neutron-star magnetosphere, it is hard to understand why spontaneous reconnection in the magnetosphere should be so slow as to produce a burst lasting many seconds, so it seems more likely that the reconnection accompanies a slower MHD process which must therefore involve a dense plasma.

It is natural to suppose that the process responsible for stressing the magnetosphere is external to the star. For instance, the stressing could be caused by the sudden infall of material due possibly to a tidally disrupted comet or to an instability of an accretion disk. However, one should bear in mind also the possibility that the stressing might arise from a process internal to the star. It is known from studies of radio pulsars that the magnetic field of a neutron star slowly decays²⁰. This requires a loss of flux from the interior and from the crust. This loss of flux might conceivably occur episodically. For instance, if the internal magnetic field is sufficiently strong and the crust sufficiently weak, the accumulation of dense magnetic flux just below the crust may on occasion lead to the fracture of the crust and the sudden injection of new flux into the magnetosphere. If this indeed occurs, a γ -ray burst may be the detectable signal of a 'magnetic volcano' on a neutron star.

I acknowledge with thanks valuable conversations with Spiro Antiochos, Fred Lamb, Edison Liang, and Vahè Petrosian. This work was supported by NASA Grants NGL 05-020-272 and NGR 05-020-668.

Received 18 November 1985; accepted 24 February 1986.

1. Woosley, S. E. (ed.) *High Energy Transients in Astrophysics* (American Institute of Physics, New York, 1984).
2. Colgate, S. A., Petschek, A. G. & Sarracino, R. in *High-Energy Transients in Astrophysics* (ed. Woosley, S. E.) 548-554 (American Institute of Physics, New York, 1984).
3. Lamb, D. Q. in *High Energy Transients in Astrophysics* (ed. Woosley, S. E.) 512-547 (American Institute of Physics, New York, 1984).
4. Sturrock, P. A. *Astrophys. J.* **164**, 529-556 (1971).
5. Hameury, J. M., Lasota, J. P., Bonazzola, S. & Heyvaerts, J. *Astrophys. J.* **293**, 56-68.

6. Liang, E. P. in *High Energy Transients in Astrophysics* (ed. Woosley, S. E.) 597-604 (American Institute of Physics, New York, 1984).
7. Liang, E. P. & Antiochos, S. K. *Nature* **310**, 121-122 (1984).
8. Svestka, Z. *Solar Flares* (Reidel, Dordrecht, 1976).
9. Sturrock, P. A. in *Solar Flares* (ed. Sturrock, P. A.) 411-449 (Colorado University Press, Boulder, Colorado, 1980).
10. Soward, A. M. & Priest, E. R. *Phil. Trans. R. Soc. A* **284**, 369-417 (1974).
11. Barat, C. in *Positron-Electron Pairs in Astrophysics* (eds Burns, M. L., Harding, A. K. & Ramaty, R.) 54-58 (American Institute of Physics, New York, 1983).
12. Nolan, P. L., Share, G. H., Forrest, D. J., Chupp, E. L., Matz, S. & Rieger, E. in *Positron-Electron Pairs in Astrophysics* (eds Burns, M. L., Harding, A. K. & Ramaty, R.) 59-63 (American Institute of Physics, New York, 1983).
13. Nolan, P. L., Share, G. H., Matz, S., Chupp, E. L., Forrest, D. J. & Rieger, E. in *High-Energy Transients in Astrophysics* (ed. Woosley, S. E.) 399-402 (American Institute of Physics, New York, 1984).
14. Kundu, M. R. *Solar Radio Astronomy* (Interscience, New York, 1965).
15. Liang, E. P. *Nature* **313**, 202-204 (1985).
16. Schaefer, B. E. *Nature* **294**, 722-724 (1981).
17. Schaefer, B. E., Seitzer, P. & Bradt, H. V. *Astrophys. J. Lett.* **270**, L49-L52 (1983).
18. Heyvaerts, J., Priest, E. R. & Rust, D. M. *Astrophys. J.* **216**, 123-137 (1977).
19. Sturrock, P. A., Kaufman, P., Moore, R. L. & Smith, D. F. *Solar Phys.* **94**, 341-357 (1984).
20. Manchester, R. N. & Taylor, J. H. *Pulsars*, 164 (Freeman, San Francisco, 1977).

Tidal reorientation and the fracturing of Jupiter's moon Europa

Alfred S. McEwen

Department of Geology, Arizona State University, Tempe, Arizona 85287 and US Geological Survey, 2255 N. Gemini Drive, Flagstaff, Arizona 86001, USA

The most striking characteristic of Europa is the network of long linear albedo markings over the surface, suggestive of global-scale tectonic processes. Various explanations for the fractures have been proposed: freezing and expansion of an early liquid water ocean¹, planetary expansion due to dehydration of hydrated silicates², localization by weak points in the crust generated by impacts³, and a combination of stresses due to planetary volume change and tidal distortions from orbital recession and orbital eccentricity^{4,5}. Calculations by Yoder⁶ and Greenberg and Weidenschilling⁷ have shown that Europa may rotate slightly more rapidly than the synchronous rate, with a rotation period (reorientation through 360°) ranging from 20 to $>10^3$ yr if a liquid mantle is present, or up to 10^{10} yr if the satellite is essentially solid⁷. Helfenstein and Parmentier⁸ modelled the stresses due to nonsynchronous rotation, and concluded that this could explain the long fractures in part of the anti-jovian hemisphere. In this note, I present a global map of lineaments with long arc lengths ($>20^\circ$ or 550 km), and compare the lineament orientations to the tensile stress trajectories due to tidal distortions (changes in the lengths of three principal semi-axes) and to nonsynchronous rotation (longitudinal reorientation of two of the principal semi-axes). An excellent orthogonal fit to the lineaments is achieved by the stresses due to nonsynchronous rotation with the axis radial to Jupiter located 25° east of its present position. This fit suggests that nonsynchronous rotation occurred at some time in Europa's history.

The global lineaments on Europa appear to be major extensional openings, ~ 10 -50 km in width, filled with new crustal material. They are darker and redder than the terrain they transect, so are probably composed of a different material³. Water ice makes up at least 95 wt% of Europa's surface⁹, so both the bright areas and the dark bands probably consist of mostly water ice, but with a greater percentage of dark contaminants in the bands. The band widths are very uniform, and are probably intrusions confined by the fracture walls rather than extrusions onto the surface. Some of the widest bands have central bright stripes, perhaps due to late-stage injections of clean ice².

Extension fractures may form either when the principal stresses are tensile or when the principal (macroscopic) stresses are compressive¹⁰. Extension fracturing during compression may be due to local tensile stresses at flaws or heterogeneities when confining pressures are low¹⁰ or the fracturing may be caused

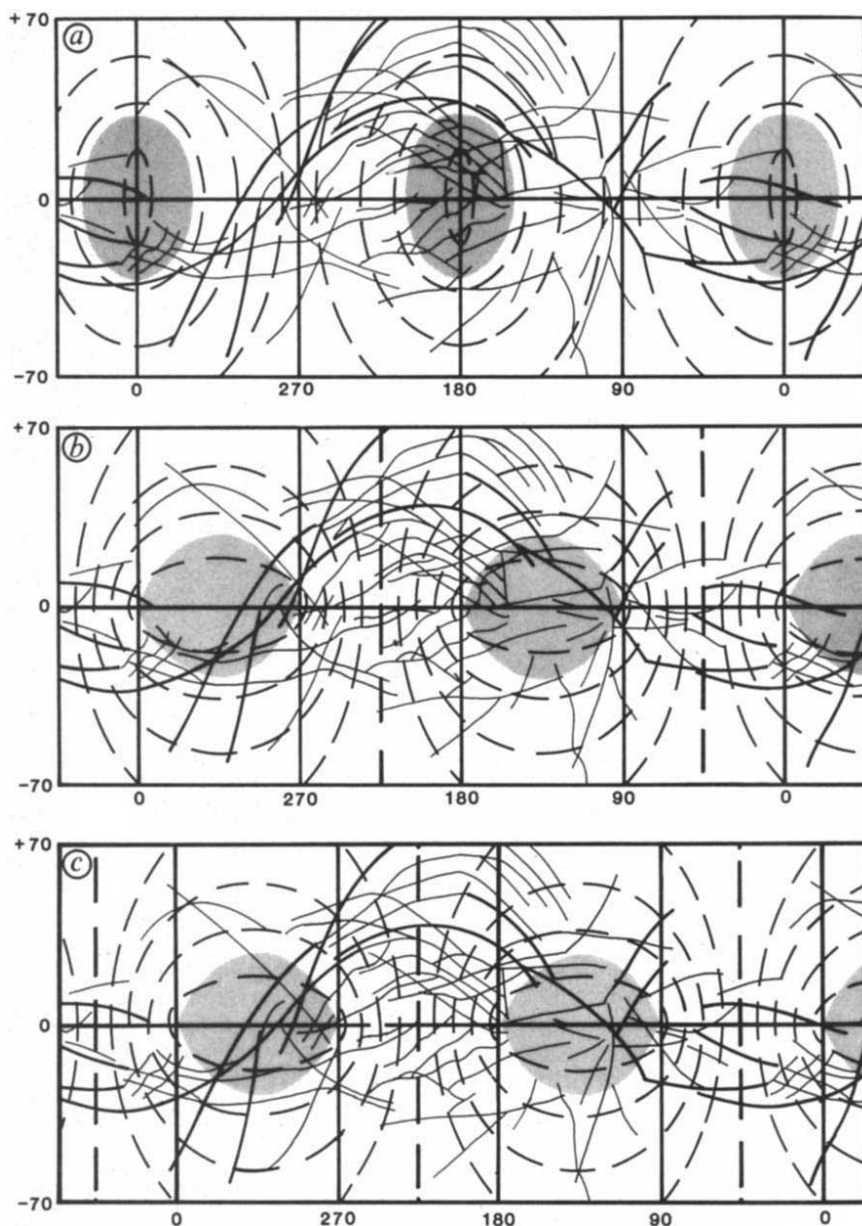


Fig. 1 Map of lineaments on Europa in Mercator projection. Heavy lines indicate great-circle lineaments. Superimposed over map (dashed lines) are trajectories for *a*, tensile and/or compressive stresses due to tidal distortions (data from ref. 5); *b*, tensile stresses due to nonsynchronous rotation with the axis radial to Jupiter at its present position (continuing nonsynchronous rotation; stress trajectories from ref. 8); and *c*, tensile stresses due to nonsynchronous rotation with the axis radial to Jupiter shifted 25° east. Shaded areas are where both principal surface stresses are compressional; applies to *a* only if trajectories are tensile.

by intrusions into the cracks resulting in forces on the crack faces¹¹. In either case, extension fractures form nearly perpendicular to the most tensile or least compressive stresses.

The best evidence for strike-slip faulting on Europa consists of minor apparent rotations near the anti-jovian region¹². A strike-slip mechanism has been proposed for some of the long global lineaments on the basis of preferred 60° orientations of lineament intersections to each other⁴ and to the tensile stresses due to tidal distortions⁵. However, there were errors in the longitudinal positions of the lineaments used in these analyses, as great as 15° (ref. 4) and 40° (refs 5, 8), so the preferred orientations are suspect. Furthermore, the tidal stresses are probably much too weak to fracture the crust in shear^{5,13}. If fractures initiate at depths where the vertical stress is the most compressive, then normal faults or graben may form¹⁴, but may become extension faults due to intrusions¹¹. Therefore, I interpret the majority of the global lineaments as extension fractures.

The lineament map, produced from Mercator projections of global mosaics of Europa¹⁵, is shown in Fig. 1. Only those lineaments visible in the global mosaics (resolution 11–23 km per pixel) and with arc lengths of at least 20° are mapped. Images utilized in the mosaics are acquired at low phase angles (3–32°),

so all of the lineaments are identified by albedo markings rather than topography. Positions are accurate to within 5°, determined by comparison with geodetic control¹⁶. Lineaments are uniformly distributed over Europa's surface, although they appear concentrated at low latitudes in Mercator projections.

Superimposed on the lineament map are the stress trajectories for tidal distortion of a thin elastic shell (Fig. 1*a*), nonsynchronous rotation with the axis radial to Jupiter at its present position (Fig. 1*b*) and nonsynchronous rotation with the axis radial to Jupiter located 25° east of the present position (Fig. 1*c*). The tidal stress trajectories (Fig. 1*a*) can represent either (1) tensile stresses due to orbital recession, or (2) compressive stresses due to orbital contraction, or (3) cyclical tensile and compressive stresses due to orbital eccentricity. Stress trajectories for nonsynchronous rotation (Fig. 1*b,c*) are tensile. The principal surface stresses are tangential (most tensile or least compressive) and normal (least tensile or most compressive) to the tensile trajectories; the stresses are reversed for compressive trajectories. Three regions of the surface may be defined by the signs of the surface stresses (++ , +− , and −− , with tensile stresses positive). Extension fractures (in all three regions) should be concentrated at a 90° angle to the tensile stress trajectories, but are least likely to form in the areas of compress-

ive (---) surface stresses, so these regions (shaded in Fig. 1) are excluded from the analysis of lineament orientations.

Very long lineaments oriented along great circles seem most likely to be due to global stresses, whereas the shorter, small-circle, and irregular lineaments may have been influenced by impacts or crustal heterogeneities. Therefore, the approximately great-circle lineaments (those whose plane passes within 15° of the centre of the sphere) of long arc length ($>30^\circ$) are shown as heavy lines on Fig. 1.

Lineament orientations to the stress trajectories were measured at intervals of 12° arc length (330 km) along each lineament. No preferred orientation with respect to tidal stresses is seen (Fig. 2a), contrary to previous reports^{4,5}, and the tidal-stress magnitudes are probably not sufficient to fracture the ice crust^{5,13}.

The lineaments are generally oriented at high angles to the tensile stress trajectories predicted for ongoing nonsynchronous rotation (Fig. 2b), consistent with extension fractures. However, if the satellite (or its crust) is continually rotating at a rate faster than synchronous, as expected⁷ if a liquid water mantle is present¹⁷, then the stresses would continuously change and fractures of different ages would not follow the same stress pattern. This can be simulated by moving the stress trajectories in Fig. 1b along the equator (with the lineaments fixed); only fractures that lie along the equator are consistent with all positions. The position giving the best fit (most orthogonal) occurs with the axis radial to Jupiter located $\sim 25^\circ$ east (relative to the surface) of its current position (Figs 1c and 2c). The 25° east position may represent the mean of stresses from reorientation through $\sim 50^\circ$ to its present position. This suggests that the fracturing resulted from an epoch of nonsynchronous rotation resulting in ~ 25 – 50° eastward reorientation of the satellite or its crust relative to the radial to Jupiter. Several additional observations support this hypothesis: (1) the great-circle lineaments give an especially good fit to the stresses (Fig. 2c); (2) a region of isotropic stress near the anti-jovian point¹⁸ may be explained by stretching this region over the tidal bulge⁷; (3) many lineaments terminate or break up into discontinuous spots³ in the regions of compressive surface stresses; and (4) the predicted stress magnitude due to a 25° reorientation is 2–3 times greater than the probable tensile strength of the ice crust⁸.

An eastward reorientation of the satellite, preserving the equator, is the direction of movement that would result from faster-than-synchronous rotation. Therefore, the proposed reorientation seems best explained by the mechanism described by Yoder⁶ and Greenberg and Weidenschilling⁷, that is, orbital eccentricity forced by resonant interaction with Io and Ganymede, resulting in non-zero tidal torque on a Europa in synchronous rotation. An alternative hypothesis is reorientation due to a large impact¹⁹, but this seems unlikely for Europa because (1) no large impact scar is apparent on the surface; (2) a 25 – 50° reorientation, preserving the equator, is improbable by this mechanism; and (3) the global fracture pattern is not like that expected from a large impact²⁰. Stable synchronous rotation is expected if the satellite has a perfectly circular orbit (not true of Europa) or if the satellite has a permanent mass asymmetry (like the Earth's Moon). The remarkable absence of topographic relief on Europa¹ is consistent with the absence of a mass asymmetry.

One possible explanation for the fracture pattern is continuous nonsynchronous rotation but with episodic fracturing when stresses have built up to exceed the tensile strength of the crust. However, this model requires some mechanism to synchronize the failure of many separate fractures. Another possibility is that venting of water vapour is very rapid¹⁷, resulting in complete resurfacing by condensed frost in a fraction of the rotation period, so that only very recent fractures ($\sim 10^3$ yr or less)^{7,8} are seen. However, the observed impact craters imply an age of at least 10^7 yr (refs 1, 3), Europa's photometric function does not support the contention that a very low-density (condensed) frost

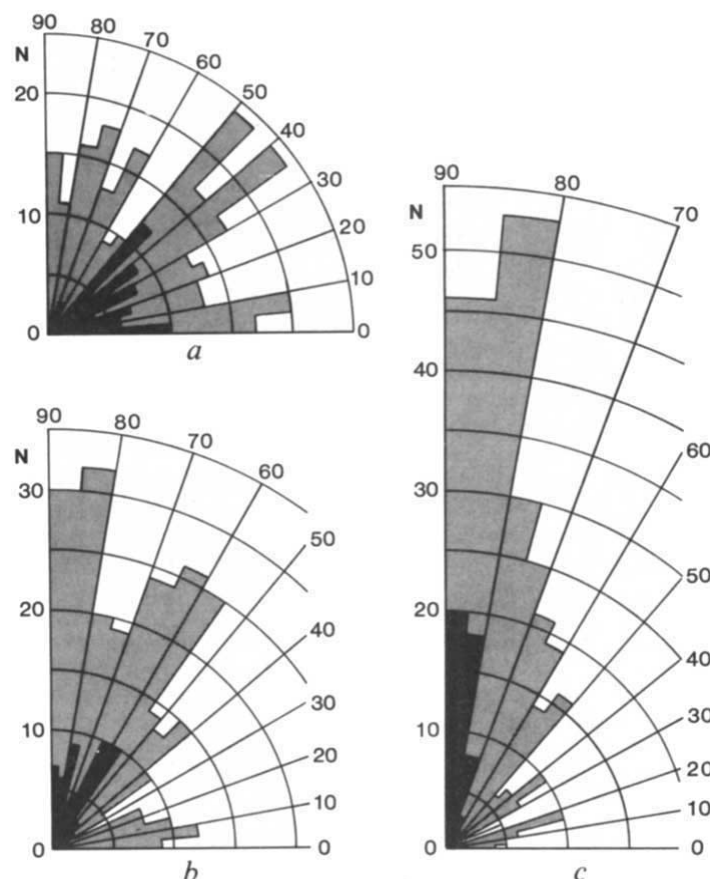


Fig. 2 Rose diagrams for angles of intersection between stress trajectories and lineaments. Black areas are for great-circle lineaments. a, b and c correspond to stress trajectories in Fig. 1.

layer is present²¹, and a thin ice shell of variable thickness¹⁷ seems unlikely to form such long, linear fractures.

Assuming that the satellite is essentially solid, there are several alternative scenarios: (1) its figure may gradually creep to a new configuration over time, resulting in very slow nonsynchronous rotation⁷, (2) an asymmetry may be dynamically supported by solid-state convection², with slowly changing convection patterns leading to epochs of nonsynchronous rotation separated by periods of synchronous rotation, or (3) the fracturing may be due to the final stages of nonsynchronous rotation before a mass asymmetry was frozen into place.

I thank R. Greenberg, E. M. Shoemaker, M. C. Malin, P. Helfenstein, B. K. Lucchitta, and J. M. Moore for critical reviews. This research was supported by NASA Planetary Geology grant NAGW-1 to Arizona State University.

Received 7 October 1985; accepted 7 March 1986.

- Smith, B. A. *et al. Science* **204**, 951–972 (1979).
- Finnerty, A. A., Ronsford, G. A., Pieri, D. C. & Collerson, K. D. *Nature* **289**, 24–27 (1981).
- Lucchitta, B. K. & Soderblom, L. A. in *Satellites of Jupiter* (ed. Morrison, D.) 521–555 (University of Arizona Press, Tucson, 1982).
- Helfenstein, P. & Parmentier, E. M. *Proc. 11th Lunar planet. Sci. Conf.*, 1987–1998 (1980).
- Helfenstein, P. & Parmentier, E. M. *Icarus* **53**, 415–430 (1983).
- Yoder, C. F. *Nature* **279**, 767–770 (1979).
- Greenberg, R. & Weidenschilling, S. J. *Icarus* **58**, 186–196 (1984).
- Helfenstein, P. & Parmentier, E. M. *Icarus* **61**, 175–184 (1985).
- Clark, R. N. *Icarus* **44**, 388–409 (1981).
- Griggs, D. T. & Handin, J. *Mem. geol. Soc. Am.* **79**, 347–364 (1960).
- Brace, W. F. in *State of Stress in the Earth's Crust* (ed. Judd, W. R.), 111–174 (Elsevier, New York, 1964).
- Schenk, P. M. *NASA TM-86247*, 3–111 (1984).
- Cassen, P. M., Peale, S. J. & Reynolds, R. T. in *Satellites of Jupiter* (ed. Morrison, D.) 93–128 (University of Arizona Press, Tucson, 1982).
- Golombek, M. P. *J. geophys. Res.* **90**, 3065–3074 (1985).
- McEwen, A. S. *J. geophys. Res.* (in the press).
- Davies, M. E. & Katayama, F. Y. *J. geophys. Res.* **86**, 8635–8657 (1981).
- Squires, S. W., Reynolds, R. T., Cassen, P. M. & Peale, S. J. *Nature* **301**, 225–226 (1983).
- Pieri, D. C. *Nature* **289**, 17–21 (1981).
- Melosh, H. J. *Earth Planet. Sci. Lett.* **26**, 353–360 (1975).
- Fleitout, L. & Thomas, P. G. *Earth planet. Sci. Lett.* **58**, 104–115 (1982).
- Buratti, B. J. *Icarus* **61**, 208–217 (1985).

Ionic structure in aqueous electrolyte solution by the difference method of X-ray diffraction

N. T. Skipper, S. Cummings*, G. W. Neilson† & J. E. Enderby*

H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK

Neutron diffraction difference methods based on isotopic substitutions offer a means of directly probing the detailed structure around a solute dissolved in a host solvent¹, but there are two main drawbacks to their implementation. First, they rely on the use of separated isotopes which are expensive and sometimes of limited availability. Second, the experiments have to be performed at a sophisticated central facility to obtain sufficiently intense beams of thermal neutrons. We have therefore initiated a series of experiments aimed at introducing difference methods to X-ray diffraction studies of liquid and amorphous systems. Difference methods have been attempted before², but the absence of definitive neutron data did not allow the method to be made fully quantitative. We now present the first results obtained from such a study in which the difference technique is applied to a concentrated aqueous electrolyte solution and the cation-cation distribution function $g_{MM}(r)$ is determined directly. The success of the method rests on the ability to identify nearly isomorphous pairs of elements, whose isomorphism can be quantified by reference to similar results obtained from the first-order difference method of neutron diffraction³.

For the case of a salt MX dissolved in water, there are ten partial structure factors, $S_{\alpha\beta}(k)$ which contribute in varying degrees to the total structure factor $F(k)$ (ref. 1). However, the (first-order) difference, $\Delta_M(k)$, between the total structure factors for any two chemically identical but isotopically distinct solutions can be written in terms of a linear combination of only the four correlations associated with the isotopically substituted ion species¹. Furthermore, the second-order difference⁴ between three such solutions gives information directly about the partial structure factor of the substituted species $S_{\alpha\alpha}(k)$. Fourier transformation of $\Delta_M(k)$ and $S_{\alpha\alpha}(k)$ gives information regarding the structural properties of the ion-water subsystem and the ions themselves through the radial pair distribution functions:

$$g_{\alpha\beta}(r) - 1 = \frac{1}{2\pi^2 nr} \int dk (S_{\alpha\beta}(k) - 1) k \sin kr$$

where n is the total number density of the system and is typically $0.1 \text{ \AA}^{-3}$.

The electrolytes chosen for this study were nickel chloride and magnesium chloride, and X-ray diffraction experiments were carried out on the following three aqueous solutions (all 4.2 M): NiCl_2 in H_2O (I), MgCl_2 in H_2O (II) and a 50:50 mixture of MgCl_2 and NiCl_2 in H_2O (III).

The X-ray structure factors, $F(k)$, of the three solutions were obtained from the raw intensity data by standard correction and normalization procedures¹.

By adopting the same formalism as in the neutron diffraction case⁴, but with the important difference that X-ray atomic form factors, f , are k -dependent, the first-order difference $\Delta_{\text{Ni}}(k)$ of the X-ray diffraction patterns for two isomorphous solutions can be written as:

$$\Delta_{\text{Ni}}(k) = A(S_{\text{NiO}}(k) - 1) + B(S_{\text{NiH}}(k) - 1) + C(S_{\text{NiCl}}(k) - 1) + D(S_{\text{NiNi}}(k) - 1) \quad (1)$$

* Present addresses: Shuster Laboratory, University of Manchester, M13 7PL, UK (S.C.); Institute Laue Langevin, BP 156X, 38042 Grenoble Cedex, France (J.E.E.).

† To whom correspondence should be addressed.

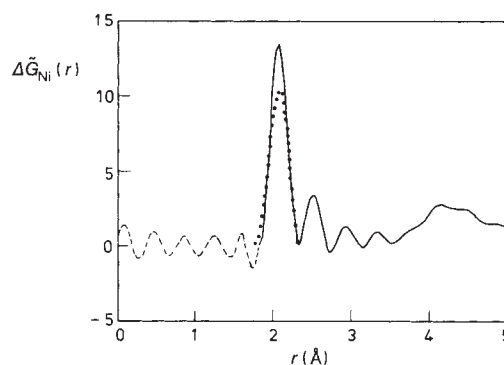


Fig. 1 The nickel ion total radial distribution function in 4.2 M aqueous chloride solution, derived from X-ray diffraction. The oscillations (dashed line) at low r are caused by termination errors in the Fourier transformation. The dotted curve is the nearest neighbour Ni-O correlation derived from neutron diffraction.

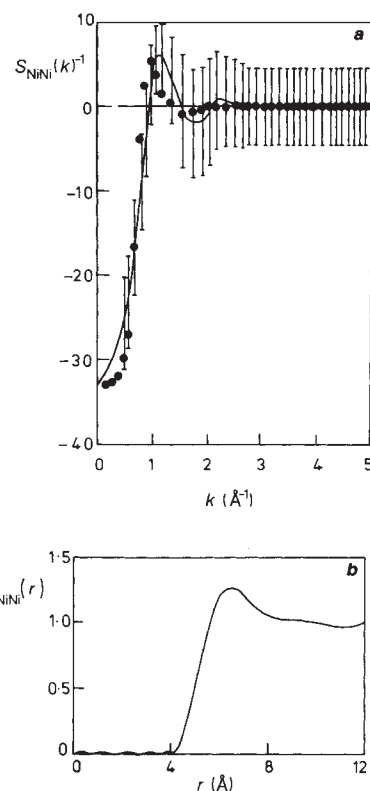


Fig. 2 a, The nickel-nickel partial structure factor, $S_{\text{NiNi}}(k)$ in 4.2 M aqueous nickel chloride solution. Curve (with error bars), X-ray diffraction; solid circles, neutron diffraction. The errors were determined by the same means as used in the neutron diffraction study⁴: a 1% systematic error was introduced into the coefficients of the $F(k)$ s. b, The corresponding Ni-Ni radial distribution function $g_{\text{NiNi}}(r)$ obtained by Fourier transformation of the curve in A after a damping function was used.

where $A = 2f_{\text{O}}(k)\Delta f_{\text{M}}(k)c_{\text{O}}c_{\text{Ni}}$, $B = 2f_{\text{H}}(k)\Delta f_{\text{M}}(k)c_{\text{H}}c_{\text{Ni}}$, $C = 2f_{\text{Cl}}(k)\Delta f_{\text{M}}(k)c_{\text{Cl}}c_{\text{Ni}}$, $D = \Delta f_{\text{M}}^2 c_{\text{Ni}}^2$, and $\Delta f_{\text{M}}(k) = (f_{\text{Ni}}(k) - f_{\text{Mg}}(k))$, $\Delta f_{\text{M}}^2 = f_{\text{Ni}}^2(k) - f_{\text{Mg}}^2(k)$, c_{β} is the concentration of species β .

The coefficients are such that A and $C \gg B, D$ so that $\Delta_{\text{Ni}}(k)$ is dominated by the nickel-oxygen and nickel chloride terms. The Fourier transformation of $\Delta_{\text{Ni}}(k)/A$, written as $\Delta\tilde{G}_{\text{Ni}}(r)$, gives $g_{\text{NiO}}(r)$ directly, together with a convolution of the other three radial distribution functions, and has the form

$$\Delta\tilde{G}_{\text{Ni}}(r) = g_{\text{NiO}}(r) + \sum_{\beta \neq \text{O}} a_{\beta} \int H_{\beta}(|\mathbf{r}-\mathbf{r}'|) g_{\text{Ni}\beta}(|\mathbf{r}'|)^{-1} d\mathbf{r}' \quad (2)$$

where $\beta = \text{H, Cl or Ni}$, $H_{\beta}(|\mathbf{r}-\mathbf{r}'|)$ is a convolution function associated with the k -dependence of the form factor and a_{β} are constants which relate to the concentration of species β .

Three $\Delta\tilde{G}_{\text{Ni}}(r)$ were calculated for the three possible differences between the $\tilde{F}(k)$ values for the three solutions studied. The largest and therefore the most accurately defined difference is

that calculated from solutions I and II, and is shown in Fig. 1. Because the first peak in $g_{\text{NiO}}(r)$ is well separated from any peaks in $g_{\text{NiCl}}(r)$, $\Delta G_{\text{Ni}}(r)$ can be used to check the isomorphism of Ni and Mg. This is achieved by comparing the result for the Ni-O nearest-neighbour correlation with that derived from the formally exact difference method of neutron diffraction. As can be seen from Table 1 and Fig. 1, the two methods yield structural parameters which are in excellent agreement. The fact that the width of $g_{\text{NiO}}(r)$ is smaller in the X-ray case reflects the superior instrument resolution obtained with our X-ray diffractometer (a θ - θ D/MAX system manufactured by Rigaku, Tokyo), relative to that of the D4 neutron diffractometer (now improved) at the Institut Laue-Langevin, even though the k -range of the scattering was about the same in both cases.

Having established that Ni and Mg are isomorphic, the cation-charge structure factor $S_{\text{NiNi}}(k)$ can be determined from a double difference⁴ between the $F(k)$ values for the three solutions:

$$S_{\text{NiNi}}(k) - 1 = \frac{1}{c_{\text{Ni}}^2(f_{\text{Mg}} - f_{\text{Ni/Mg}})} \times \left[\frac{(F_{\text{I}}(k) - F_{\text{II}}(k))}{(f_{\text{Ni}} - f_{\text{Mg}})} - \frac{(F_{\text{I}}(k) - F_{\text{III}}(k))}{(f_{\text{Ni}} - f_{\text{Ni/Mg}})} \right] \quad (3)$$

where $f_{\text{Ni/Mg}}(k) = (f_{\text{Ni}}(k) + f_{\text{Mg}}(k))/2$.

The Fourier transformation of $S_{\text{NiNi}}(k)$ yields $g_{\text{NiNi}}(r)$. The results shown in Fig. 2 are in good agreement with those obtained

Table 1 Comparison of peak position $\bar{r}_{\text{NiO}}$, coordination number $\bar{n}$ and half-width Δr for $g_{\text{NiO}}(r)$ derived from X-ray and neutron diffraction

Method	$\bar{r}_{\text{NiO}}$ (Å)	$\bar{n}$	Δr (Å)
X-ray	2.07 (2)	6.0 (2)	0.23 (2)
Neutron	2.07 (2)	5.8 (2)	0.26 (2)

from an earlier neutron diffraction experiment in which isotopic substitution was used and represent the first ion-ion radial distribution function obtained by X-ray methods alone.

The above results clearly demonstrate the power of the X-ray difference method. The proof that Ni and Mg are isomorphic in aqueous electrolyte solution means that a comprehensive study in which counter-ion, concentration and temperature are varied, can now be undertaken. For example, it will be possible to obtain accurate information regarding the ion-oxygen distribution at a much lower concentration than is currently practicable with the neutron difference method. Experiments on NiBr_2 and MgBr_2 aqueous solutions will help resolve the controversy regarding the existence of inner-sphere complexing in bromide solution.

The X-ray difference method may play an increasingly important role in the determination of the structural properties of a variety of liquid and amorphous systems, and a systematic procedure is being developed in order to identify other isomorphic pairs.

We thank SERC for their continuing support of the aqueous solution programme, and Mr Philip Gullidge for assistance in the calibration and preparation of samples. J.E.E. thanks NATO for the award of a travel grant (RG125/80) which allowed some of the ideas contained in this paper to be discussed with Professor Harold Friedman at SUNY, Stony Brook.

2 September 1985; accepted 13 February 1986.

Constant-geometry edge-flaking of brittle materials

E. A. Almond & N. J. McCormick

National Physical Laboratory, Teddington, Middlesex TW11 0LW, UK

Lithic tool manufacture was the earliest technological exploitation of the brittle fracture of hard materials, and archaeologists have traced and reproduced the evolution of flake tools through various stages of development from the early simple flake knives (Fig. 1) to the sophisticated multi-faceted edges of stone axes. The technology advanced during the period of gunflint manufacture. Of special interest are the simple geometrical relationships between the dimensions of flakes formed in certain controlled conditions^{1,2}. The phenomenon of edge-flaking is also encountered in the attrition of tool cutting edges and in the chipping of precision edges of metrological components, where it is avoided rather than understood, by empirical radiusing (rounding-off). In our effort to improve the efficiency of cutting edges, we have observed that a law of constant geometry seems to apply to the flakes formed at rectangular edges, and the effect is relatively insensitive to the test material.

Our experimental approach has been to indent rectangular blocks of the test material with increasing loads at different distances from an edge until a flake breaks off. The distance at which flaking occurs is approximately equal to the flake thickness (Fig. 2). We used a conical diamond indenter with a tip radius of 0.2 mm and an apex angle of 120°. We initially observed that the size of the flakes formed in a tungsten carbide/cobalt tool material decreased with decrease in distance from the edge, and the width-to-height and thickness-to-height ratios of the flakes were approximately constant (Fig. 3). When the same test procedure was extended to other materials it was found that similar height-to-width and thickness-to-width ratios applied for flakes formed in various glasses, a high-carbon steel, an alumina ceramic, a Thames Valley flint (Fig. 4a), a polymethyl methacrylate polymer and a range of tungsten carbide/cobalt tool materials. Furthermore, when the boxed volume (defined as the product of the height, width and thickness of the flake) was plotted against the true volume (calculated from the weight and

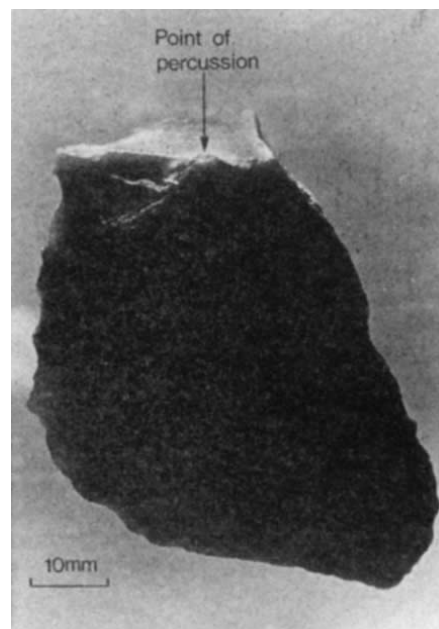


Fig. 1 Heavily struck flake from lower gravels, Barnfield, Swanscombe, Kent (UK); Clactonian culture, 300,000 yr old.

1. Enderby, J. E. & Neilson, G. W. *Rep. Prog. Phys.* **44**, 593-653 (1981).
2. Bol, W., Gerrits, G. J. A. & Eck, C. L. vP. *J. appl. Crystallogr.* **3**, 486-492 (1970).
3. Soper, A. K., Neilson, G. W., Enderby, J. E. & Howe, R. A. *J. Phys. C10*, 1793-1801 (1977).
4. Neilson, G. W. & Enderby, J. E. *Proc. R. Soc. A390*, 353-371 (1985).
5. Habenschuss, A. & Spedding, F. H. *J. chem. Phys.* **70**, 2797-2806 (1979).

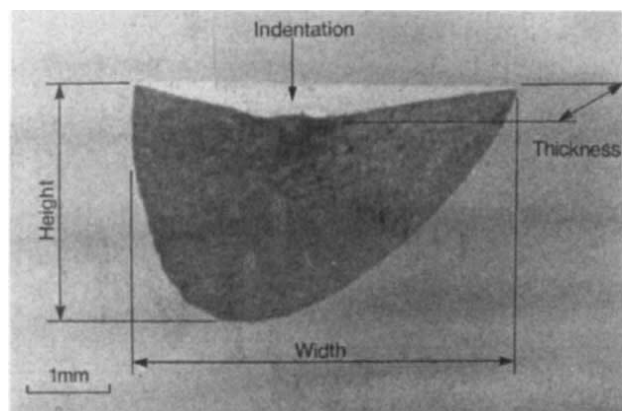


Fig. 2 Flake produced by static loading of the edge of a tungsten carbide/cobalt tool materials.

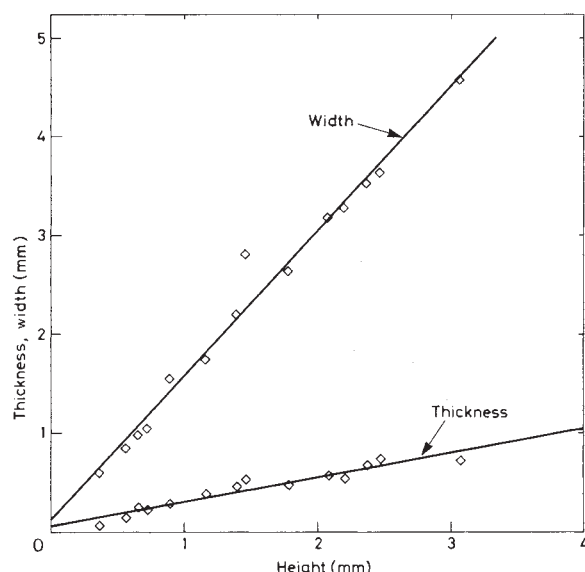


Fig. 3 Variation of flake width and thickness with flake height for a 9 wt% cobalt tungsten carbide tool material; correlation coefficients, $r = 0.9901$ and 0.9564 , respectively. (The correlation coefficient r is defined as $r = \sum(x - \bar{x})(y - \bar{y}) / (\sum(x - \bar{x})^2 \sum(y - \bar{y})^2)^{1/2}$, where x, y and $\bar{x}, \bar{y}$ are respectively the coordinates of a data point and the corresponding point on the best straight line obtained by minimizing the sum of the squares of the residuals.)

density of the flake) a linear relationship was observed (Fig. 4b), with a slope of ~ 0.25 for all the materials tested.

Although detailed inspection of the data in Fig. 4 indicates that they may belong to separate curves according to the size of the test piece, manufacturing technique and hardness, the observed linear relationship has unexpected implications. Thus, for edge flakes produced by statically loaded indenters: (1) the volume lost from the edge depends only on the distance of the loading point from the edge and is relatively independent of the material; and (2) the shape of the flake varies little; it is scaled in proportion to the distance from the edge.

Examination of previous data³ showed that the observations of constant geometry also held for flakes produced by ball indenters of 2–5 mm diameter using tungsten carbide/cobalt tool materials for the indenter and the test materials. However, in the earlier work the measurement techniques were not sufficiently sensitive to detect the flake size dependence on distance from the edge.

The existence of a constant-geometry law for the dimensional ratios implies that a similar law might apply for the angle of inclination between the flat and curved vertical faces of the

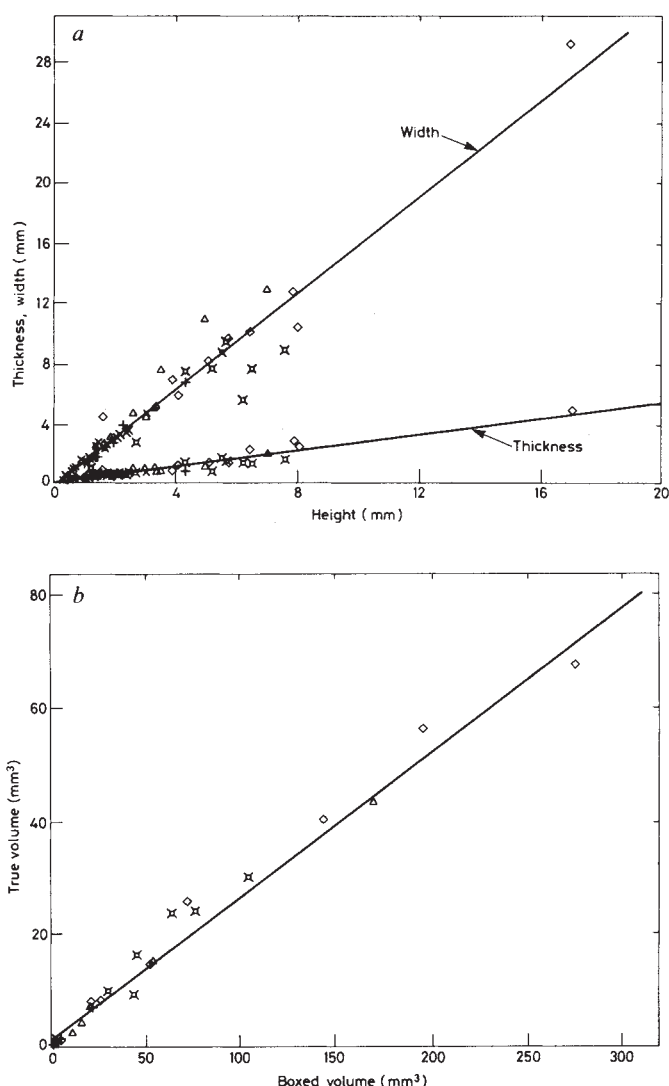


Fig. 4 a, Variation of flake width and thickness with flake height ($r = 0.9685$ and 0.9579 , respectively); b, relation between flake volume and boxed volume ($r = 0.9997$). Data for various glasses ($\square$), Thames Valley flint ($\diamond$), an alumina ceramic ($\triangle$), a 9 wt% cobalt tungsten carbide tool material ($\times$) and a high-carbon steel ($+$); b also contains results from ref. 3.

flake, and for the angle subtended by the flake faces at the indentation. However, although the latter tended towards 146° , (as expected from the ratio of 20:3 for the width and thickness dimensions), it was not possible to develop a consistent measurement criterion because the faces were convex and the curvatures varied considerably. This applied even more to the angle of inclination, because the flake surface was also convex in the thickness direction and the flakes were not normally symmetric about a vertical plane through the indentation, as can be seen in Fig. 2.

There must be a materials dependence of the flaking phenomenon, and this is found in the magnitude of the load required to initiate flaking at a given distance from the edge, which varies, for example, from 164 N for a silica glass to 6.15 kN for a 9 wt% cobalt tungsten carbide tool material at a distance of 0.71 mm from an edge. The mechanism of flake formation is not yet understood, but it is not restricted to fracture nucleation by hertzian cracking and conical fracturing as has been suggested for flakes produced in lithic tools with blunt hammers in percussive conditions^{1,2}. In our tests, although circumferential crack paths of hertzian origin have been observed at some of the

fracture initiation sites, the fracture path more commonly passes through the centre of the indentation, which is characteristic of the tensile stress field resulting from plastic deformation or densification under the loading point⁴. It would probably be more appropriate to consider flaking as two energy-dissipative processes: a load-dependent crack initiation mechanism which occurs in the deformed region below the indenter, and a crack propagation stage determined by the strain energy release rate in linear elastic fracture mechanics conditions⁵.

The fact that good reproducibility of flake geometry was obtained with different indenter shapes and different materials may be related to the quasi-static loading mode used in our tests. Thus, stone-age humans adopted pressure loading 25,000 yr ago to dress the edges of cutting tools⁶, and it is tempting to conclude that this technique was favoured because it was more precise and reproducible than the percussive loading methods developed by their ancestors.

We acknowledge valuable discussions with Dr Mark Newcomer of the Institute of Archaeology, London, and the loan from his collection of the tool shown in Fig. 1.

Received 31 December 1985; accepted 26 February 1986.

1. Speth, J. D. *Am. Antiq.* **37**, 34–60 (1972).
2. Speth, J. D. *Am. Antiq.* **40**, 203–207 (1975).
3. Almond, E. A. & Roebuck, B. *Proc. int. Conf. Science of Hard Materials* (eds Viswanadham, R. K. *et al.*) 597–614 (Plenum, New York, 1983).
4. Evans, A. G. & Wilshaw, T. R. *Acta metall.* **24**, 939–956 (1976).
5. Fonseca, J. G., Eshelby, J. D. & Atkinson, C. *Int. J. Fracture Mech.* **7**, 421–433 (1971).
6. Clark, G. & Pigot, S. *Prehistoric Societies* (Penguin, Harmondsworth, 1970).
7. Sachs, L. *Applied Statistics* (Springer, New York, 1982).

Inferred oxygen isotope profile of Archaean oceanic crust, Onverwacht Group, South Africa

S. E. Hoffman*, M. Wilson† & D. S. Stakes‡

* College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA

† Department of Geological Sciences, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0

‡ Department of Geology, University of South Carolina, Columbia, South Carolina 29208, USA

Whole-rock oxygen isotope data from a suite of mafic and ultramafic samples from the Onverwacht Group, the basal unit of the Archaean greenstone succession of the Barberton Mountain Land, South Africa, show a range of values from $\delta^{18}\text{O} \approx +3$ to $\delta^{18}\text{O} \approx +14$, a range which coincides with those of Phanerozoic ophiolites and oceanic crust. When the samples are arranged in an inferred ophiolitic pseudostratigraphy, from basal serpentinitized ultramafic cumulates through altered mafic extrusives, they have an oxygen isotope distribution profile which is indistinguishable from that of Phanerozoic ophiolites. The distinctive isotopic profiles and secondary mineral assemblages in Phanerozoic ophiolites are caused by hydrothermal interaction between seawater ($\delta^{18}\text{O} \approx 0$) and oceanic crust ($\delta^{18}\text{O} \approx +5.8$). The existence of a similar isotopic and metamorphic profile in the Onverwacht Group argues strongly that these rocks were hydrothermally altered by an Archaean ocean whose isotopic composition was indistinguishable from modern sea water.

The Barberton Mountain Land is one of the oldest surviving Archaean greenstone belts. Whole-rock Sm–Nd measurements have given the oldest ages: 3.54 ± 0.03 Gyr (ref. 1) and 3.56 ± 0.24 Gyr (ref. 2). Other isotopic systems have given ages clustered about 3.3 Gyr (refs 3–8).

The rocks of the greenstone belt are divided stratigraphically into the Moodies, Fig Tree, and Onverwacht Groups. The Onverwacht Group, in the stratigraphy of Viljoen and Viljoen⁹, encom-

Table 1 Multiple runs of rose quartz standard (RQ) and duplicate runs of selected Onverwacht samples

Sample ID	$\delta^{18}\text{O}$ (SMOW)	Average	Plotted
RQ-CIT	+8.48		—
RQ-CIT	+8.49		—
RQ-CIT	+8.50	+8.45	—
RQ-CIT	+8.18		—
RQ-CIT	+8.83		—
RQ-CIT	+8.24		—
RQ-US	+8.26		—
RQ-US	+8.29		—
RQ-US	+8.36	+8.30	—
RQ-US	+8.27	($2\sigma < 0.1$)	—
RQ-US	+8.31		—
RA5052(US)	+4.58	+4.52	+4.5
RA5052(US)	+4.45		
RA5054(US)	+5.47	+5.27	+5.3
RA5054(US)	+5.06		
RA5055(US)	+6.47	+6.56	+6.6
RA5055(US)	+6.65		
RA5063(US)	+5.20	+5.24	+5.2
RA5063(US)	+5.28		
RA13175(US)	+2.45	+2.58	+2.6
RA13175(US)	+2.70		

CIT, California Institute of Technology; US, University of Saskatchewan. Gregory and Taylor²⁰ reported a $\delta^{18}\text{O}$ value for RQ of +8.45‰ obtained at CIT. Analytical reproducibility for the data plotted in Fig. 1 is calculated as $2\sigma < 0.3\%$.

passes six formations of mafic and ultramafic extrusive and intrusive rocks which have been altered to chloritic greenschists, actinolitic amphibolites, and serpentinites, and which include minor intercalations of felsic volcanics and volcanoclastic and carbonate sediments^{9,10}. The lithologies of each formation are described in detail by Viljoen and Viljoen⁹ and by Tankard *et al.*¹⁰. The upper subgroup, the Geluk, also known as the Mafic to Felsic Unit, is composed of the Swartkoppie, Kromberg, and Hoogenoeg Formations⁹. The lower subgroup, the Tjakastad, also known as the Lower Ultramafic Unit, is composed of the Komati, Theespruit, and Sandspruit Formations⁹. In order to account for what they described as repetitious cycles of volcanics and sediments up to 20 km in total thickness, Viljoen and Viljoen¹¹ proposed a model of cyclical volcanism and sedimentation.

More recent mapping in the Barberton Mountain Land^{12–15} has produced evidence that the stratigraphic succession may not be one of simple supercession. Williams and Furnell¹² recognized structural unconformities within the Onverwacht Group which suggested isoclinal folding and faulting, and De Wit¹³ proposed that thrust faulting had occurred, resulting in tectonic repetition of strata. These studies^{12–15} suggest that the 20-km thickness is a tectonic artefact rather than the result of volcanological or sedimentological processes.

De Wit¹⁴ has compiled a revised stratigraphy which consists of the layered ultramafic complexes at the base, progressively overlain by the Komati Formation, the Hoogenoeg and Kromberg Formations, the rhyodacites and silicic porphyries of the Mafic to Felsic Unit, the Fig Tree Group, and the Moodies Group. De Wit and Stern¹⁶ previously proposed that the Onverwacht Group constitutes an Archaean ophiolite sequence. In this paper, we report an additional parallel between the Onverwacht Group and Phanerozoic ophiolites and oceanic crust: a whole-rock oxygen isotope distribution which reflects cooling and alteration by convective seawater hydrothermal activity.

A total of thirty-three powdered 'whole-rock' samples from rocks collected in the Komati Formation, the Hoogenoeg Formation, and the Stolzberg Ultramafic Complex by Roger Hart and

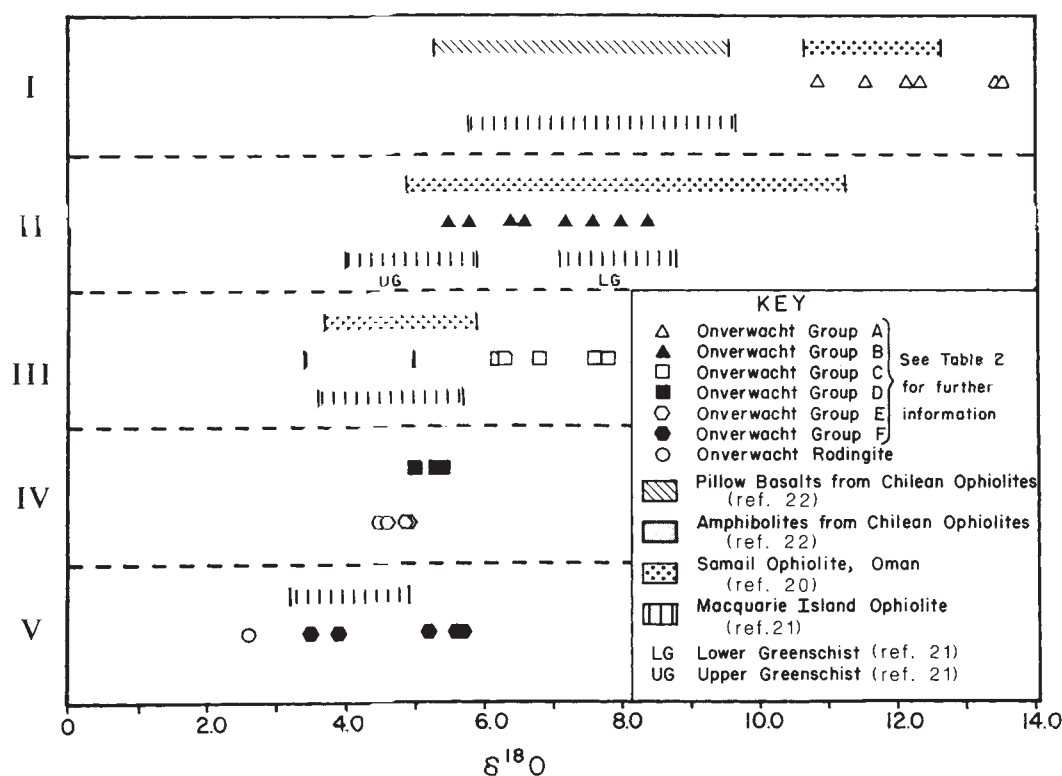


Fig. 1 Whole-rock $\delta^{18}\text{O}$ distribution in the Onverwacht Group plotted in an ophiolitic pseudostratigraphy (size of symbols is larger than range of calculated experimental error), along with data from the Oman ophiolite²⁰, the Macquarie Island ophiolite²¹, and Chilean ophiolites²². The Onverwacht data are compared with the following lithologies from the Phanerozoic ophiolites: I, pillow basalts; II, sheeted dykes; III, gabbros; IV, (only Onverwacht data shown); V, serpentinitized harzburgites.

Maarten De Wit were analysed for their bulk oxygen isotope composition at the California Institute of Technology and at the University of Saskatchewan. Oxygen was liberated using the fluorine technique¹⁷ at CIT, and the bromine pentafluoride technique¹⁸ at the University of Saskatchewan. Data are reported relative to standard mean ocean water (SMOW, $\delta^{18}\text{O} = 0\text{‰}$). On this scale, NBS-28 (quartz) has a $\delta^{18}\text{O}$ value of $+9.6\text{‰}$ ¹⁹. Statistical data on analytical precision and accuracy as well as interlaboratory comparisons are shown in Table 1.

The data from sixteen whole-rock analyses at CIT and 32 whole-rock analyses at the University of Saskatchewan are listed in Table 2. The samples are grouped according to their alteration mineralogies and relict textures. The results of the 32 analyses from the University of Saskatchewan and the analysis of RA13177 from CIT are shown in Fig. 1. The Onverwacht rocks are plotted according to an ophiolitic pseudostratigraphy which is derived in part from the revised stratigraphic model of De Wit¹⁴. The pseudostratigraphy proceeds 'upsection' from ultramafic rocks (Groups D, E, and F in Table 2, and IV and V in Fig. 1) through progressively less mafic rocks to altered pillow basalts (Groups A, B, and C in Table 2 and I, II, and III in Fig. 1). Figure 1 also shows ranges of $\delta^{18}\text{O}$ for corresponding samples from the Samail ophiolite in the Sultanate of Oman²⁰, the Macquarie Island ophiolite²¹, and ophiolites of Southern Chile²². There is a rough correlation between $\delta^{18}\text{O}$ and lithology for the samples from Oman, Macquarie Island, Chile, and the Onverwacht Group, although there are also significant overlaps between fields. The range of whole-rock $\delta^{18}\text{O}$ values relative to SMOW for the Onverwacht Group ($\delta^{18}\text{O} = +2.6$ – 13.6‰) coincides almost entirely with the combined range of the Macquarie Island, Samail, and Chilean ophiolites ($\delta^{18}\text{O} = +3.2$ – 12.7‰). If the rodingite from the Stolzberg Ultramafic Complex is excluded (RA13175, Table 2), the ranges completely coincide except for two Onverwacht samples, RA5061 and RA5062, which are dominantly composed of smectite²³, a mineral which is enriched in ^{18}O (ref. 24).

Chemical data from the samples listed in Table 2²³ show that the concentrations of all major oxides and most minor elements are distinctly correlated with oxygen isotope value, water content, and CO_2 content²³. This indicates that chemical alteration

of the rocks was coincident with the redistribution of their oxygen isotope content.

Basaltic rocks altered in subaerial hydrothermal systems by meteoric water have substantially different oxygen isotope signatures from those altered in submarine systems by seawater because meteoric water is depleted in ^{18}O relative to seawater. This depletion results from isotopic fractionation during evaporation and increases with increasing latitude²⁵. The average whole-rock $\delta^{18}\text{O}$ value for rocks altered in meteoric hydrothermal systems is significantly lower than that of rocks altered by seawater at the same temperatures of alteration²⁵. That is, metamorphic assemblages of the same grade will be much more depleted in ^{18}O , and thus have lower $\delta^{18}\text{O}$ values, in a meteoric hydrothermal system. For example, greenschist-facies rocks altered in meteoric hydrothermal systems at temperatures sufficient for the formation of actinolite and chlorite have $\delta^{18}\text{O}$ values relative to SMOW between roughly -2 and $+4$ (ref. 25). In contrast, oxygen isotope studies of Phanerozoic submarine greenstones have revealed that their $\delta^{18}\text{O}$ values range from $+2.8$ to $>+10$ (refs 24–26). Taking into account the analytical accuracy of the study, Fig. 1 shows that the Onverwacht rocks fall totally within the combined fields for Phanerozoic oceanic crust and ophiolites ($\delta^{18}\text{O} \approx +2.8$ to $>+10$)^{20–22,24–26}.

This coincidence of isotopic and lithologic characteristics between samples from the Onverwacht Group and those from Phanerozoic ophiolites and oceanic crust could indicate that the Onverwacht rocks were altered by Archaean sea water in an oceanic environment. If one assumes that the $\delta^{18}\text{O}$ value of the Archaean ocean was essentially equivalent to that of the present ($\approx 0\text{‰}$), then it is possible that the same conditions and processes which produced the oxygen isotope distribution in the Samail²⁰, Macquarie Island²¹, and Chilean²² ophiolites were also responsible for the oxygen isotope distribution observed in samples from the Onverwacht.

There have been relatively few stable isotope studies on samples from Archaean greenstone belts, but there is some evidence that the $\delta^{18}\text{O}$ value of the ocean has changed very little over the past 3 Gyr. Beatty²⁷ analysed more than 150 samples from six areas within the Abitibi greenstone belt (2.8 Gyr). In each of the six areas, he found an increase in $\delta^{18}\text{O}$ with strati-

Table 2 Whole-rock oxygen isotope values for samples from the Onverwacht Group

	$\delta^{18}\text{O}$	$\delta^{18}\text{O}$		$\delta^{18}\text{O}$	$\delta^{18}\text{O}$
Sample	Caltech	U. Sask.	Sample	Caltech	U. Sask.
A Clay-calcite			D Spinifex-textured komatiites*		
RA5060	—	+12.4	RA5054	+5.3	+5.3
RA5061	—	+13.6	RA13174	—	+5.0
RA5062	—	+13.5	RA13259	—	+5.4
RA13256	—	+10.9			
RA13257	—	+12.2			
RA13258	—	+11.6			
B Chlorite-calcite-(epidote)-(qtz)			E Serp. dunites, Komati Formation		
RA5055	+6.1	+6.6	RA5052	—	+4.5
RA5056	+6.0	+5.8	RA5053	+4.5	+4.9
RA5057	+6.1	+5.5	RA13178	+5.5	+4.9
RA13173	+7.0	+7.6	RA13181	—	+4.6
RA13180	+7.2	+7.2			
RA13182	—	+6.4			
RA13227	+8.1	+8.4			
RA13228	—	+8.0			
C Actinolite-albite-epidote			F Samples from ultramafic complexes†		
RA5048	+7.2	+7.6	RA5063	+5.1	+5.2
RA5049	—	+6.2	RA5068	—	+3.9
RA5050	+6.8	+6.8	RA5069	—	+3.5
RA13172	+6.8	+6.3	RA13175	+2.7	+2.6
RA13260	—	+7.7	RA13176	+6.0	+5.6
			RA13177	+5.8	—

Samples are divided into groups according to their alteration mineralogies and their relict igneous textures. Results from California Institute of Technology were obtained in 1982 using the fluorine extraction method¹⁷. Results from the University of Saskatchewan were obtained in 1984 using the bromine pentafluoride method¹⁸. The average difference between analyses from the two different laboratories is 0.3‰, with $2\sigma < 0.5\%$.

* Mineralogy of spinifex-textured komatiites: spinifex blades now altered to serpentine (and magnetite); interstices between blades now altered to tremolite and magnetite.

† RA5063 and RA13176 are partially altered orthopyroxenites; RA5068 and RA5069 are serpentinized dunites; RA13175 is a rodingite with a probable orthopyroxenite precursor; RA13177 is a massive serpentine whose precursor is not determinable.

graphic height which correlated with the hydrothermal alteration of the sequence, and in five of the six areas, he found that the altering fluid was probably sea water with a $\delta^{18}\text{O}$ value between -2 and $+2$ (ref. 27). Thus, the data of Beatty²⁷ and Beatty and Taylor^{28,36} support the assumption that the $\delta^{18}\text{O}$ of sea water has varied little for at least the past 3 Gyr.

The Onverwacht oxygen isotope data also strongly imply that the range of temperatures over which submarine hydrothermal alteration occurs has not markedly changed over the past 3.4 Gyr. The range of temperatures at which rocks are altered in a submarine hydrothermal system is mainly dependent upon the temperature of the magma and the tectonically controlled configuration of the seawater convection system relative to the magma chamber, not on the temperature of the intake water (that is, ambient bottom water). Thus, the Onverwacht rocks provide little data to clarify the debate on Archaean bottom water temperatures. However, it should be noted that most estimates of Archaean seawater temperatures and $\delta^{18}\text{O}$ values are based upon analyses of cherts. In a study of Phanerozoic cherts recovered from Deep Sea Drilling Project cores, Hein and Yeh²⁹ determined that the samples did not reflect palaeoenvironmental (bottom water) temperatures but rather diagenetic temperatures of formation which were generally greater than 30°C . While bottom water temperatures have fluctuated in the Cretaceous, Phanerozoic ambient bottom temperatures have been generally less than 10°C .

Based predominantly on studies of Onverwacht cherts, Perry and Tan³⁰ suggested that the $\delta^{18}\text{O}$ value of the Archaean ocean

was significantly lower than at present and that there has been a secular increase in the ^{18}O content of the oceans over the history of the Earth. However, their evaluation of their samples as having no significant post-depositional diagenetic history is suspect in view of more recent studies. The evidence for intense hydrothermal alteration, silicification, and recrystallization in Onverwacht strata is overwhelming^{6,23,31}. In addition, many Onverwacht "cherts" are in fact silicified pyroclastic deposits^{31,32}, and the silicifying fluids appear to have been evolved hydrothermal solutions resulting from the interaction between seawater and underlying mafic and ultramafic strata³¹. Thus, the $\delta^{18}\text{O}$ values of Onverwacht cherts reflect their interaction with diagenetic fluids significantly different from Archaean ambient bottom water.

On the basis of oxygen isotope measurements of oceanic crustal samples and mass balance calculations, Muehlenbachs and Clayton³³ proposed that the $\delta^{18}\text{O}$ value of the oceans has been buffered via "processes associated with the creation, aging, and destruction of the oceanic crust"³³. They calculated that at present-day seafloor creation rates, the oceans would achieve a $\delta^{18}\text{O}$ value of ~ 0 roughly 250 Myr after the onset of seafloor spreading³³. If seafloor creation rates were five times the present day values³⁴, this 'time constant' would equal only 60 Myr (ref. 33).

Abbott and Hoffman³⁴ proposed a model for the evolution of plate tectonic processes as a function of the aging of oceanic crust over time. This model postulates that seafloor creation rates 3.8 Gyr ago were as much as five times present-day values. If the time constant calculated for this seafloor creation rate is indeed 60 Myr (ref. 33), then the oxygen isotope composition of the oceans would have been buffered by hydrothermal alteration before the formation of the Onverwacht Group. Therefore, the oxygen isotope distribution in the Onverwacht could be the result of the same types of processes which Gregory and Taylor²⁰ and Cocker *et al.*²¹ proposed to explain the distributions observed in the Samail and Macquarie Island ophiolites.

In such a system³⁵, sea water with $\delta^{18}\text{O} \approx 0\%$ interacts with ultramafic and gabbroic rocks ($\sim +5.7$ – 6.0%) at depth, altering their mineralogies and isotopically exchanging oxygen (and hydrogen) at temperatures above 500°C over a range of mostly low (<1) water/rock ratios. The interaction leaches ^{18}O from the rocks, enriching the fluids in ^{18}O . These ^{18}O -enriched fluids percolate upward into overlying strata where they alter those rocks to greenschist mineralogical assemblages at somewhat lower temperatures. The interaction of these rocks with ^{18}O -enriched fluids at lower temperatures causes them to be relatively enriched in ^{18}O . The high positive $\delta^{18}\text{O}$ values of the overlying basalts result from low-temperature seafloor alteration (weathering). Stratigraphic preservation in the Onverwacht may not permit the kind of detailed modelling of the structure of the hydrothermal system as was performed by Gregory and Taylor²⁰ and Cocker *et al.*²¹, particularly with regard to the effects of magma chamber configuration on the fluid flow regime. However, the general principle of oxygen isotope redistribution under conditions of subseafloor hydrothermal metamorphism is undoubtedly applicable.

Detailed descriptive information on the samples analysed in this study may be obtained by writing to S.E.H. We thank R. Hart for permission to analyse rock samples from his Onverwacht collection and Drs H. P. Taylor Jr and K. Kyser for access to their laboratories. Thanks also to K. Kyser, E. Nesbit, E. Suess, R. Duncan, M. Lyle, D. Abbott and M. De Wit, all of whom reviewed earlier drafts of the manuscript and substantially improved it through their comments. S.E.H. was funded by grant no. DPP-8120473 from the NSF Division of Polar Programs. M.W. and laboratory expenses at the University of Saskatchewan were funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to T. K. Kyser. D.S.S. was at the California Institute of Technology during part of this study. Isotopic analyses at CIT were funded

by a grant from NSF Division of Ocean Sciences, OCE80-19021, to Dr H. P. Taylor Jr.

Received 18 March 1985; accepted 28 January 1986.

- Hamilton, P. J. *et al.* *Nature* **279**, 298–300 (1979).
- Jahn, B. M., Gruau, G. & Glikson, A. Y. *Contr. Miner. Petrol.* **80**, 25–40 (1982).
- Jahn, B. M. & Shih, C. *Geochim. cosmochim. Acta* **38**, 611–627 (1974).
- Van Niekirk, C. B. & Burger, A. J. *Trans. geol. Soc. S. Afr.* **72**, 9–25 (1969).
- Hurley, P. M., Nagy, B. & Pinson, W. H. Jr *Earth planet. Sci. Lett.* **14**, 360 (1972).
- De Wit, M. J. *et al.* *Econ. Geol.* **77**, 1783–1802 (1982).
- York, D. *et al.* *Eos* **62**, 429 (1981).
- Martinez, M. L. *et al.* *Nature* **307**, 352–354 (1984).
- Viljoen, M. J. & Viljoen, R. P. *Spec. Pub. geol. Soc. S. Afr.* **2**, 9–151 (1969).
- Tankard, A. J. *et al.* *Crustal Evolution of Southern Africa* (Springer, New York, 1982).
- Viljoen, M. J. & Viljoen, R. P. in *African Magmatism and Tectonics* (eds Clifford, T. N. & Gass, I. G.) 27–49 (Oliver and Boyd, Edinburgh, 1970).
- Williams, D. & Furnell, R. G. *Precamb. Res.* **9**, 325–347 (1979).
- De Wit, M. J. *J. struct. Geol.* **4**, 117–136 (1982).
- De Wit, M. J. *Spec. Pub. geol. Soc. S. Afr.* **9**, 185–187 (1983).
- De Wit, M. J., Fripp, R. E. P. & Stanistreet, I. G. *Spec. Pub. geol. Soc. S. Afr.* **9**, 21–29 (1983).
- De Wit, M. J. & Stern, C. R. *2nd int. Archaeol. Symp., Perth* 85–87 (Geological Society of Australia, 1980).
- Taylor, H. P. Jr & Epstein, S. *Geol. Soc. Am. Bull.* **73**, 461–480 (1962).
- Clayton, R. N. & Mayeda, T. K. *Geochim. cosmochim. Acta* **27**, 43–52 (1963).
- Coplen, T. B., Kendall, C. & Hoppie, J. *Nature* **302**, 236–238 (1983).
- Gregory, R. T. & Taylor, H. P. Jr *J. geophys. Res.* **86**(B4), 2737–2755 (1981).
- Cocker, J. D., Griffin, B. J. & Muehlenbachs, K. *Earth planet. Sci. Lett.* **61**, 112–122 (1976).
- Stern, C. R., De Wit, M. J. & Lawrence, J. R. *J. geophys. Res.* **81**, 4370–4380 (1976).
- Hoffman, S. E. thesis, Oregon State Univ. College of Oceanography (1985).
- Stakes, D. S. & O'Neil, J. R. *Earth planet. Sci. Lett.* **57**, 285–304 (1982).
- Taylor, H. P. Jr in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A. *et al.*) 83–139 (Plenum, New York, 1983).
- Muehlenbachs, K. & Clayton, R. N. *Can. J. Earth Sci.* **9**, 471–478 (1972).
- Beatty, D. W. thesis, California Institute of Technology (1980).
- Beatty, D. W. & Taylor, H. P. Jr *Econ. Geol.* **77**, 95–108 (1982).
- Hein, J. R. & Yeh, H.-W. *Init. Rep. DSDP* **62**, 749–758 (1981).
- Perry, E. C., Jr & Tan, F. C. *Geol. Soc. Amer. Bull.* **83**, 647–664 (1972).
- Heinrichs, T. *Precamb. Res.* **24**, 237–283 (1984).
- Lowe, D. R. & Knauth, L. P. *J. Geol.* **85**, 699–723 (1977).
- Muehlenbachs, K. & Clayton, R. N. *J. geophys. Res.* **81**, 4365–4369 (1976).
- Abbott, D. H. & Hoffman, S. E. *Tectonics* **3**, 439–448 (1984).
- Taylor, H. P. Jr *J. geol. Soc. Lond.* **133**, 509–558 (1977).
- Beatty, D. W. & Taylor, H. P. Jr in *Komatites* (ed Arndt, N. T. & Nisbet, E. G.), 267–280 (Allen & Unwin, London, 1982).

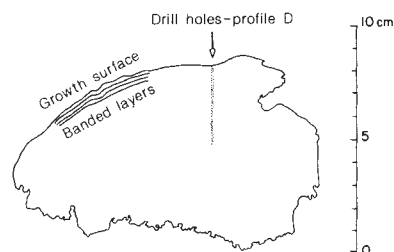


Fig. 1 Schematic diagram showing the sampling intervals for stable isotopes along profile D (0.5-mm drill holes) and for ^{210}Pb and ^{14}C analyses (0.5–1.0 mm bands)¹⁹. The $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ measurements obtained from the profile with the highest resolution are discussed in the text. Data obtained from three other profiles are not reported because an unexpected ageing effect after drilling was observed for these samples. Although it is standard practice to store aragonitic CaCO_3 samples indefinitely, we found that the $\delta^{18}\text{O}$ values of these drilled samples were lower by 0.3–0.4‰ after storage periods of a few months; most of the $\delta^{13}\text{C}$ results were decreased by 0.1–0.15‰. When samples were stored for a period of months after baking this effect was even greater. The effect of baking was observed previously by Land *et al.*¹⁷, but these workers did not mention storage effects. We suspect that high-speed drilling of dense aragonite heats up the material sufficiently to cause partial conversion to calcite, a mineral that can exchange its CO_3^{2-} with other sources of CO_2 . The results reported here are for samples that were analysed within 48 h of drilling, thus ensuring the smallest possible ageing effect.

Input of excess CO_2 to the surface ocean based on $^{13}\text{C}/^{12}\text{C}$ ratios in a banded Jamaican sclerosponge

Ellen R. M. Druffel & Livia M. Benavides*†

Department of Chemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

* Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The CO_2 content of the atmosphere has increased during the past two centuries as a result of the combustion of fossil fuels for energy¹ and the reduction of forest and soil carbon reservoirs on land². The amount of CO_2 added to the atmosphere from fossil-fuel burning is known from historical records¹ ($\pm 10\%$), but the contribution from reduction of the terrestrial biosphere is far less certain. Several authors have estimated the relative contributions from the two sources by measuring the change in the $^{13}\text{C}/^{12}\text{C}$ ratio in atmospheric CO_2 as revealed in tree rings^{3–8} (CO_2 derived from these two sources is depleted in ^{13}C with respect to that in the atmosphere). Using trees in the Northern Hemisphere, recent estimates of the integrated CO_2 release from the terrestrial biosphere since AD 1800 ranged from 70% (ref. 5) to 90% (ref. 9) of that released from fossil fuels. Here we present surface ocean $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ records measured in the skeleton of a living sclerosponge (*Ceratoporella nicholsoni*), which accretes aragonite in isotopic equilibrium with the surrounding sea water/dissolved inorganic carbon (DIC) system. The $\delta^{13}\text{C}$ record reveals a decrease

of 0.50‰ from 1820 to 1972. Using a model of the world carbon cycle and a deconvolution of our $\delta^{13}\text{C}$ data, we estimate that the amount of excess CO_2 derived from the terrestrial biosphere is ~38% of that from fossil-fuel sources. Our model calculations support a preindustrial CO_2 concentration in the atmosphere of 280 p.p.m.v. (parts per million by volume), in agreement with direct measurements of air occluded in Antarctic ice cores¹⁰.

Tree-ring records may not accurately reflect changes in atmospheric $\delta^{13}\text{C}$ with time because of kinetic isotope fractionation during the diffusion of CO_2 through the leaf stomata and during carboxylation of ribulose diphosphate during photosynthesis^{11,12}. Small variations of the fractionation associated with these and other^{11,12} processes may alter the tree-ring $\delta^{13}\text{C}$ record, rendering it unrepresentative of the actual change in the atmosphere due to excess CO_2 input.

Another approach for determining the relative contributions of fossil-fuel and forest/soil CO_2 is to examine $\delta^{13}\text{C}$ records obtained from the surface waters of the oceans. Nozaki *et al.*¹³ found a reduction in $\delta^{13}\text{C}$ of 0.50‰ from 1900 to 1974 in a banded coral (*Diploria*) from Bermuda. However, other authors have found changes in coral $\delta^{13}\text{C}$ for this period that are too large to be attributed solely to the input of excess CO_2 from the atmosphere (ref. 14 and E.R.M.D., unpublished data). Corals contain symbiotic algae (zooxanthellae) whose photosynthetic activities fractionate the DIC pool from which the coral accretes its aragonite. This and other processes constitute the 'vital effect' which, as in trees, might vary with time masking the actual $\delta^{13}\text{C}$ change in sea water. Isotopic studies of corals devoid of symbiotic algae reveal a $\delta^{13}\text{C}$ record more representative of the ambient DIC in sea water, as that part of the vital effect associated with photosynthesis is absent^{15,16}. However, Land *et al.*¹⁷ found a direct correlation between the $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ in several shallow-water ahermatypic corals, indicating a common effect on both stable isotope ratios. Land *et al.* proposed that this was caused by variable retention rates of metabolic CO_2 in the internal DIC pool, another process contributing to the vital effect.

In the present study, we used the sclerosponge *Ceratoporella nicholsoni* as a recorder of $^{13}\text{C}/^{12}\text{C}$ and $^{18}\text{O}/^{16}\text{O}$ ratios in sea water. This animal has no symbiotic algae and appears to precipitate aragonite without a vital effect. The sclerosponge was

† Present address: Avenue Pardo 239, Apto 1310, Lima 18, Peru.

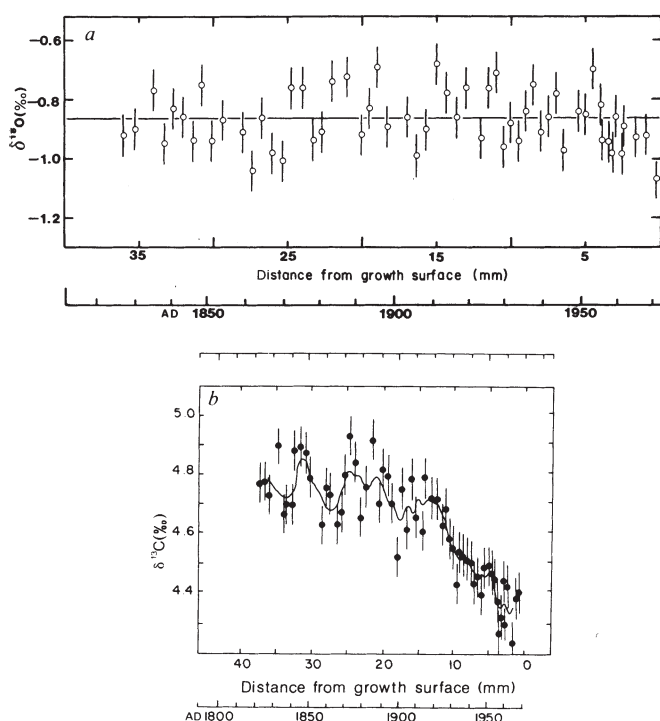


Fig. 2 *a*, $\delta^{18}\text{O}$ measurements for drilled samples along profile D. There is no apparent trend in these data, suggesting that there was no significant change in the long-term averaged SST, salinity and water composition from 1820 to 1972 on the reef. The solid line represents a least-squares fit of the data. Error bars represent 1σ error of $\pm 0.07\%$, determined from numerous triplicate analyses. *b*, $\delta^{13}\text{C}$ measurements for drilled samples along profile D. A decrease of 0.50% from the mid-1800s to 1972 is apparent. The solid line represents a cubic spline function fitted to the data (with 4-point smoothing). The error bars represent a 1σ error of $\pm 0.07\%$.

collected in 1972 in the fore-reef zone of Chalet Caribe Reef from the north-west coast of Jamaica at a depth of 26 m. This area is laved by open ocean waters. The major current system influencing this western Caribbean location is the precursor to the Gulf Stream.

The sclerosponge (~7 cm in diameter) was cut into slabs 3–4 mm wide which were studied using a dissecting microscope to ensure that they had been cut parallel to the growth axis. The slabs were cleaned ultrasonically to eliminate remnants of worm tubes and dust from cutting. A thin slab (1 mm) was X-rayed in an attempt to discern growth bands of varying density; however, despite the use of high-contrast film, no bands were detected. Dark-light band pairs, approximately 1 mm wide, were visible to the naked eye. These bands were identified by Hartman and Goreau¹⁸ as a part of the organic matrix of the sclerosponge. The slabs were mapped using these organic bands (a total of 26) and sectioned for radioacarbon and ^{210}Pb analyses¹⁹. To obtain higher resolution of the stable isotope records, samples were drilled along a profile on a single slab (see Fig. 1). The profiles were drilled at intervals of 0.5 to 1.0-mm. For stable isotope analyses, the samples were baked at 375°C for 1 h and converted to CO_2 by acidification with 100% orthophosphoric acid under vacuum at 50°C . A VG Micromass 602E mass spectrometer was used to make the measurements.

The growth rate of this Jamaican sclerosponge was determined¹⁹ using ^{210}Pb and bomb ^{14}C distributions in growth layers near the active growing surface. From these measurements and from the pre-bomb ^{14}C distributions in the sclerosponge bands, the specimen was estimated to have lived from ~AD 1700 to 1972; the average growth rate was $0.25 \pm 0.05 \text{ mm yr}^{-1}$ using the two methods. This gives an uncertainty of $\pm 30 \text{ yr}$ for our oldest year assignment (AD 1825).

Figure 2*a* and *b* shows the $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ results, plotted against distance from the growing surface and also against date of formation based on the bomb ^{14}C chronology¹⁹.

The $\delta^{18}\text{O}$ measurements in Fig. 2*a* show no obvious trends. The entire data set gives an average $\delta^{18}\text{O}$ value of -0.87% (s.d. = 0.07%). These data suggest that there was no significant long term variation in any of the factors controlling the $^{18}\text{O}/^{16}\text{O}$ ratio of the accreted aragonite^{20–25} (that is, sea surface temperature (SST), salinity or seawater composition). Variations in $\delta^{18}\text{O}$ caused by seasonal SST changes are insignificant, as each sample represents a 2.5-yr growth interval. Surface salinity measurements were made along a transect extending from the shore to the collection site, which was ~600 m offshore. These results reveal that salinity values at the collection site (36.055–36.059‰) were within the range of values present in the western Caribbean Sea. The slight salinity gradient away from shore (on the order of $+0.03\%$) may represent a small input of fresh water from the island. A variation of this magnitude, however, would not significantly affect the $\delta^{18}\text{O}$ values ($\pm 0.015\%$).

A cubic spline function was fitted to the $\delta^{13}\text{C}$ data set, with 4-point smoothing (solid line in Fig. 2*b*). $\delta^{13}\text{C}$ values from sclerosponge layers accreted during the nineteenth century average $+4.78 \pm 0.09\%$ (1 s.d.). Superimposed on this average value is an inter-annual oscillation of 0.2% with a timescale of approximately two decades. A trend toward values that are 0.40 – 0.50% lower is evident in the upper 15 mm of the sclerosponge skeleton. This result is similar to that of Nozaki *et al.*¹³, who revealed a decline in $\delta^{13}\text{C}$ of 0.50% in a Bermuda coral. Although there is little correlation between the short-term variations of $\delta^{13}\text{C}$ as recorded in the surface ocean by the sclerosponge and the Bermuda coral (Fig. 3), partially because of the uncertainty in the chronology of the sclerosponge bands, the overall trend toward lower values by 1972 is the same.

There is no apparent correlation between the $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values in samples from the sclerosponge, in contrast to the direct relationship found in ahermatypic corals¹⁷. Thus, the two quantities are independent, which is essential if we are to conclude that the isotopic ratios in the skeletal material represent accretion in equilibrium with the seawater/carbonate system.

The $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values expected for equilibrium inorganic precipitation of calcite in surface waters of the northwestern Atlantic Ocean are -1.4 to -2.3% (at SST of $28 \pm 1^\circ\text{C}$) and 2.4 to 2.6% , respectively^{17,26–28}. To compare these ranges with our data, we had to take into account the additional fractionation that occurs during the accretion of aragonite as opposed to calcite, which is $+0.6\%$ for $\delta^{18}\text{O}$ (ref. 29) and 1.8% for $\delta^{13}\text{C}$ (ref. 28). Thus, values for aragonite precipitated recently in equilibrium with surrounding sea water in the northern Jamaican region are -0.8 to -1.7% for $\delta^{18}\text{O}$ and 4.2 to 4.4% for $\delta^{13}\text{C}$. Our results fall within these ranges (average $\delta^{18}\text{O}$ of -0.87% , $\delta^{13}\text{C}$ range of 4.2 – 4.5%). This evidence, in addition to the fact that there is no apparent correlation between $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$, suggests that the vital effect is insignificant and that the sclerosponge precipitates aragonite in equilibrium with the DIC/water

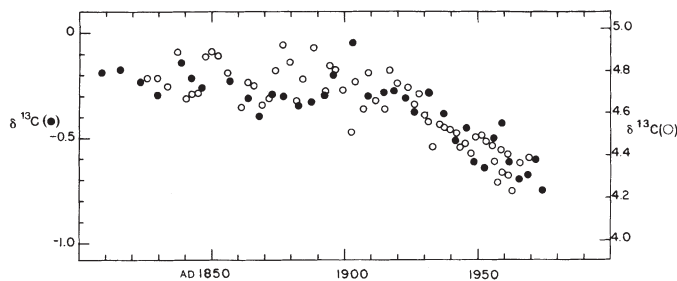


Fig. 3 $\delta^{13}\text{C}$ trend in surface ocean water in the north-west Atlantic as revealed by a Bermuda coral¹³ (●, 5-yr averages) and by a Jamaican sclerosponge (present study) (○, 2.5-yr averages). A decrease of $0.50 \pm 0.10\%$ is apparent from the mid-1800s to 1972.

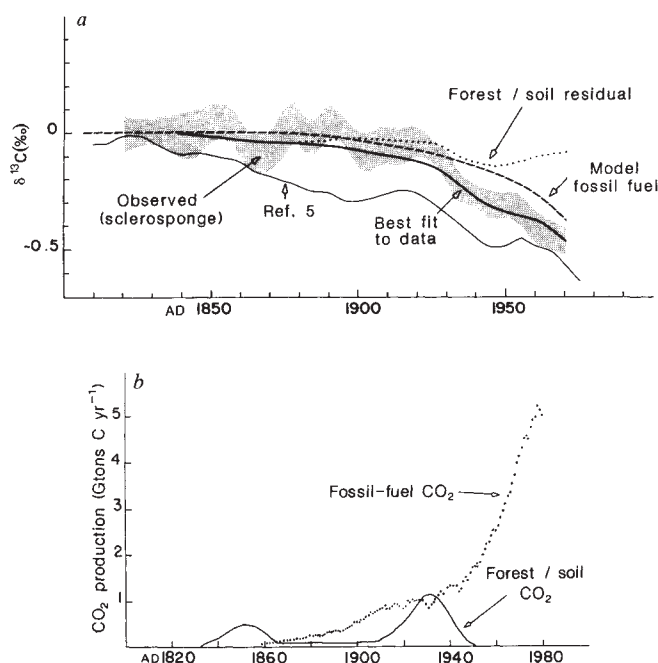


Fig. 4 *a*, Model calculations of the $\delta^{13}\text{C}$ trend in surface ocean waters. The stippled area represents the observed values from the sclerosponge. The decrease due to fossil-fuel CO_2 alone (---) was calculated from historical records^{1,33}. The forest/soil component (....) was calculated by eliminating the fossil-fuel curve from the model-derived best fit of the observed sclerosponge measurements (heavy solid line). The surface-ocean curve calculated from the tree-ring data of Stuiver *et al.*⁵ is shown for comparison (light solid line). *b*, Annual CO_2 production from fossil-fuel burning¹ and from forest/soil sources (present work) during the past two centuries. Integration of the two curves reveals that ~28% of the total excess CO_2 released into the atmosphere during 1820–1972 was derived from the terrestrial biota.

system in which it lives. In contrast to corals, sclerosponges pump large amounts of water through their channels¹⁸; this may provide a mechanism for the removal of metabolic CO_2 from the animal tissue, thus eliminating significant retention of respired CO_2 in the calcioblastic layer, and hence in the accreted aragonite.

Assuming that sclerosponges accrete aragonite in equilibrium with the surrounding sea water, their skeletons contain valuable information regarding the input of excess CO_2 to the oceans. Knowledge of both the timing and magnitude of the $\delta^{13}\text{C}$ change obtained from the stable isotope records in banded sclerosponge layers enables us to estimate the relative contributions to the atmosphere and oceans of the two major sources of CO_2 : fossil fuels and forest/soil carbon.

A one-dimensional mixing model of the world carbon cycle, similar to that of Emanuel *et al.*³⁰, was used to estimate the timing and magnitude of excess CO_2 input to the atmosphere from the two major sources mentioned above. In this model, the ocean is constructed of 19 layers that are mixed diffusively, in a manner similar to the box-diffusion model of Oeschger *et al.*³¹. The biota is represented by the five major terrestrial carbon reservoirs. Values of the atmospheric P_{CO_2} (partial pressure of CO_2) calculated using the model agreed (to within ± 2 p.p.m.v.) with measurements obtained by Keeling *et al.*³² from 1958 to 1982 at Mauna Loa.

The stippled region in Fig. 4*a* shows the range of $\delta^{13}\text{C}$ values in the surface ocean as recorded in the sclerosponge bands. Using the model, we calculate that the reduction in $\delta^{13}\text{C}$ expected in average surface ocean waters due to the input of fossil CO_2 alone^{1,33} is 0.37‰ by 1972 (broken line); this is two-thirds of the value observed in the sclerosponge record. If

we assume that the north-west Atlantic is representative of average global surface ocean water, then it seems that there is an additional source of excess CO_2 , presumably from deforestation and soil manipulation on land. In the model, we have added CO_2 derived from the reduction of the forest/soil reservoir ($\delta^{13}\text{C} = -26\text{‰}$) in order to obtain the best fit of the observed ocean $\delta^{13}\text{C}$ curve (heavy solid line in Fig. 4*a*). The change in $\delta^{13}\text{C}$ in the surface ocean calculated from the composite tree-ring curve of Stuiver *et al.*⁵ is slightly greater than our observed trend (Fig. 4*a*). Figure 4*b* shows the annual CO_2 production (in gigatons C yr^{-1}) from the two sources of excess CO_2 . The forest/soil contribution is approximately 38% the size of the fossil fuel contribution, representing a net input to the atmosphere of 53 G tons C from the terrestrial biosphere between 1820 and 1972. Using the combined observed ocean trend (Fig. 3) from the Bermuda coral and Jamaican sclerosponge, the estimate of the forest/soil CO_2 contribution remains the same.

Although the shape of the forest/soil curve is difficult to determine with any certainty because of the large signal-to-noise ratio in the ocean $\delta^{13}\text{C}$ records, these data do suggest that the change in $\delta^{13}\text{C}$ due to forest/soil CO_2 input was minimal between 1950 and 1972; this is in agreement with the results of studies of tree rings^{5,9} and atmospheric CO_2 (ref. 34). Also, there was a significant decrease in $\delta^{13}\text{C}$ from 1845 to 1865, suggesting that the forest/soil component was an important net source of CO_2 to the atmosphere during the nineteenth century. However, it appears from the sclerosponge data that most of the excess terrestrial CO_2 was expelled into the atmosphere during the period 1920–1950 (Fig. 4*b*).

The bimodal nature of the forest/soil CO_2 input, as revealed here, has also been demonstrated by deconvolution of ice-core CO_2 data³⁵ and of tree-ring $\delta^{13}\text{C}$ records^{5,36}. The similarity of recent results derived from independent geochemical data sets indicates that our understanding of the excess CO_2 problem is approaching a consensus—(1) the net input of CO_2 to the atmosphere from the biosphere between 1800 and 1975 was approximately half of that from fossil fuels, and (2) the net production of biospheric CO_2 has undergone two periods of peak production in the past two centuries, at 1820–1850 and 1880–1930.

An estimate of the partial pressure of CO_2 in the atmosphere during the early 1800s can be made using these data. Using historical records of the industrial production of CO_2 from fossil fuels^{1,33} and the P_{CO_2} record in the atmosphere³², the preindustrial P_{CO_2} is estimated at 292 p.p.m.v., if the biosphere was neither a sink nor a source of CO_2 . Considering the biotic source of excess CO_2 , which is estimated here to have been ~38% the size of the fossil-fuel source, the preindustrial P_{CO_2} is decreased to 280 p.p.m.v.; this value is in agreement with measurements made by Neftel *et al.*¹⁰ of air trapped in ice cores from Siple Station, Antarctica.

We thank Peter Williams, Peter Brewer, Phil Dustan, Dave Keeling, Lloyd Keigwin, Ann McNichol, Barrien Moore, Bill Emanuel and Werner Deuser for helpful discussions; Phil Dustan for providing the sclerosponge; Sheila Griffin for technical support; Debbie Shafer for helping with the modelling efforts; Stephen Rodriguez for collecting the salinity samples; and Molly Lumping for typing the manuscript. This research was supported by NSF grants OCE81-11954 and OCE83-15260.

Received 11 November 1985; accepted 27 February 1986.

1. Keeling, C. D. *Tellus* **25**, 174–198 (1973).
2. Houghton, R. A. *et al. Ecol. Monogr.* **53**, 235–262 (1983).
3. Rebellio, A. & Wagener, K. *Environmental Biogeochemistry* (ed. Nriagu, J. O.) 13–23 (Ann Arbor Science, Ann Arbor, 1976).
4. Pearman, G. J., Francey, R. J. & Fraser, P. J. *B. Nature* **260**, 771–773 (1976).
5. Stuiver, M., Burk, R. L. & Quay, P. D. *J. Geophys. Res.* **89**, (D7), 11731–11748 (1984).
6. Harkness, D. D. & Miller, B. F. *Radiocarbon* **22**, 291–298 (1980).
7. Francey, R. J. *Nature* **290**, 232–235 (1981).
8. Freyer, H. D. *Proc. 6th ORNL Life-Sci. Symp.*, Oak Ridge (eds Trabalka, J. R. & Reichle, D. E.) (Springer, New York, in the press).
9. Peng, T.-H. *Geophys. Monogr.* **32**, 123–131 (1985).
10. Neftel, A., Moore, E., Oeschger, H. & Stauffer, B. *Nature* **315**, 45–47 (1985).

11. Farquhar, G. D., O'Leary, M. H. & Berry, J. A. *Aust. J. Pl. Physiol.* **9**, 121-137 (1982).
12. Francey, R. J. & Farquhar, G. D. *Nature* **297**, 28-31 (1982).
13. Nozaki, Y., Rye, D. M., Turekian, K. K. & Dodge, R. E. *Geophys. Res. Lett.* **5**, 825-828 (1978).
14. Chivas, A. R., Aharon, P., Chappell, J., Vlastuin, C. & Kiss, E. *Proc. Inaugural Great Barrier Reef Conf.*, Townsville, 77-81 (J.C.U. Press, Townsville, 1983).
15. Weber, J. N. *Deep Sea Res.* **20**, 901-910 (1973).
16. Emiliani, C., Hudson, J. H., Shinn, E. A. & George, R. Y. *Science* **202**, 627-629 (1978).
17. Land, L. S., Lang, J. C. & Barnes, D. J. *Geochim. cosmochim. Acta* **41**, 169-172 (1977).
18. Hartman, W. D. & Goreau, T. F. *Symp. zool. Soc. Lond.* **25**, 205-243 (1970).
19. Benavides, L. & Druffel, E. R. M. *Coral Reefs* (in the press).
20. Swart, P. K. *Earth Sci. Rev.* **19**, 51-80 (1983).
21. Weber, J. N. & Woodhead, P. M. J. *J. geophys. Res.* **77**, 463-473 (1972).
22. Fairbanks, R. G. & Dodge, R. E. *Geochim. cosmochim. Acta* **43**, 1009-1020 (1979).
23. Weil, S. M., Buddemeier, R. W., Smith, S. V. & Kroopnick, P. M. *Geochim. cosmochim. Acta* **45**, 1147-1153 (1981).
24. Dunbar, R. B. & Wellington, G. M. *Nature* **293**, 453-455 (1981).
25. Druffel, E. R. M. *Geophys. Monogr.* **32**, 111-122 (1985).
26. Kroopnick, P. *Deep Sea Res.* **32**, 57-84 (1985).
27. Epstein, S., Buchsbaum, R., Lowenstam, H. A. & Urey, H. C. *Bull. geol. Soc. Am.* **64**, 1315-1326 (1953).
28. Robinson, M. & Clayton, R. N. *Geochim. cosmochim. Acta* **33**, 997-1002 (1969).
29. Tarutani, T., Clayton, R. N. & Mayeda, T. K. *Geochim. cosmochim. Acta* **33**, 987-996 (1969).
30. Emanuel, W. R., Killough, G. G., Post, W. M., Shugart, H. H. & Stevenson, M. P. *Dep. Envir. Docum. No. DOE/NBB-0062, UC-11*, 79 (1984).
31. Oeschger, H., Siegenthaler, U., Schotterer, U. & Gugelmann, A. *Tellus* **27**, 168-192 (1975).
32. Keeling, C. D., Bacastow, R. B. & Whorf, T. P. *Carbon Dioxide Review: 1982* (ed. Clark, W. C.) 377-385 (Oxford University Press, 1982).
33. Rotty, R. M. *Carbon Cycle Modelling, SCOPE 16* (ed. Bolin, B.) 121-123 (Wiley, New York, 1981).
34. Keeling, C. D., Mook, W. G. & Tans, P. P. *Nature* **277**, 121-123 (1979).
35. Siegenthaler, U. & Oeschger, H. *Tellus* (in the press).
36. Peng, T.-H. & Freyer, H. D. *Proc. 6th ORNL Life Sci. Symp. Oak Ridge* (eds Trabalka, J. R. & Reichle, D. E.) (Springer, New York, in the press).

Molecular evidence for a terrestrial component of organic matter dissolved in ocean water

Kathleen J. Meyers-Schulte & John I. Hedges

School of Oceanography, WB-10, University of Washington, Seattle, Washington 98195, USA

Dissolved organic matter (DOM) in seawater represents one of the largest active carbon reservoirs on Earth¹. Although mass-balance calculations suggest a substantial riverine input to the marine DOM pool², a terrestrial organic component has not been positively identified in open-ocean water. By using lignin-derived phenols as molecular-level probes of DOM (analogous to previous studies in sediments³⁻⁵), we report here the first unambiguous evidence for the presence of terrestrially derived DOM in open ocean water. Dissolved humic substances, isolated by resin adsorption from near-surface water of the eastern equatorial Pacific, yield lignin-derived phenols in compositional patterns which resemble those obtained from Amazon River water^{6,7}. Total phenol yields from these open-ocean humic isolates are, on average, ~10% of those from Amazon humic substances, indicating that ~10% of dissolved marine humic material (and at least 0.5% of the bulk marine DOM) is terrestrially derived.

The concentration of terrestrially derived organic matter in the marine DOM pool is one of the most ill-defined factors in the global carbon cycle¹. The potential for a significant terrestrial contribution is evident from the calculation that the global DOM input by rivers² is sufficient to support ~70% of the oceanic DOM pool at its measured (¹⁴C) average age of 3,400 yr⁸. Evidence that riverine DOM mixes conservatively in selected estuaries⁹ suggests that most riverine DOM passes through estuaries to coastal waters¹⁰, where its subsequent fate is unknown.

Direct biochemical analyses of marine DOM have yielded limited source-related information because the overall abundance of recognizable biochemicals is low (10-20%)¹¹⁻¹³ and their geographic origins are usually ambiguous. A portion of the DOM is comprised of a complex mixture of acidic polymers called humic substances which are produced by microbial degradation and condensation of biochemicals and can be isolated by adsorption onto resin¹⁴. Bulk isotopic and chemical analyses of humic isolates from seawater indicate a predominant marine

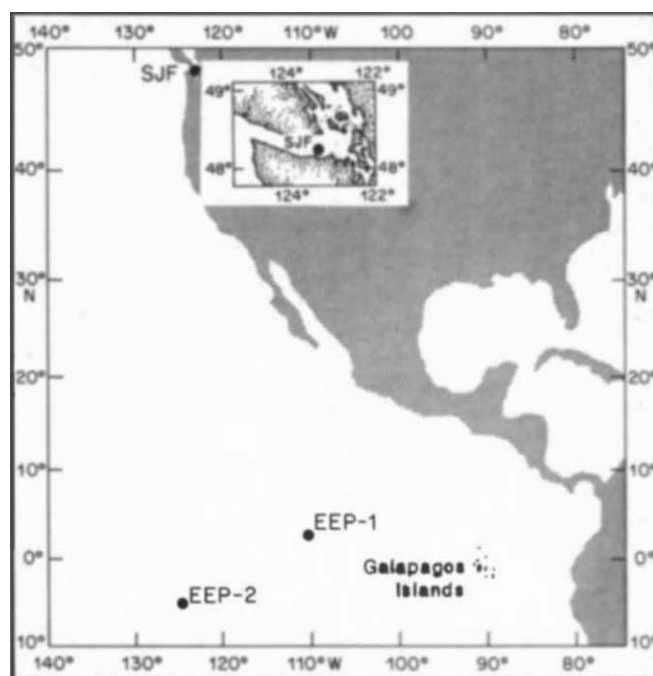


Fig. 1 Seawater sampling sites.

origin¹⁵⁻¹⁸, but these measurements are not well suited for detection of minor constituents, and cannot preclude a terrestrial component. Degradative analyses of humic substances allow the greater sensitivity inherent in molecular level analyses, but to date have either yielded products of unclear geographic origin¹⁸, have been applied to landlocked marine water¹⁰ which may not be representative of the open ocean, or have not detected a terrestrial component of marine humus¹⁹.

Using the lignin-biomarker technique⁵, we report here the first direct evidence for the presence of terrestrially derived organic matter in open-ocean DOM. Lignin is a phenolic polymer which is produced solely by vascular land plants as structural units of their cell walls²⁰. Its exclusive terrestrial source, high natural abundance on land, inherent chemical stability²¹ and unmistakable suite of cupric oxide oxidation products render lignin an attractive molecular tracer for terrestrial organic matter in aquatic environments^{6,7,22}. On oxidation by cupric oxide, lignin of all vascular plant tissues produces vanillyl phenols, which can be used as unambiguous general tracers of terrestrially derived material. In addition, syringyl phenols (produced only from angiosperm lignin) and cinnamyl phenols (produced only from non-woody tissues) denote specific vascular tissue sources^{4,22}. Finally, the weight ratio of vanillic acid to vanillin, symbolized as (Ad/Al)_v of fresh (undergraded) vascular plant tissues ranges from 0.1 to 0.2 (ref. 23), whereas higher values are characteristic of biodegraded lignin⁵.

Previous applications of the lignin-biomarker technique to dissolved humic substances from lakes and rivers have demonstrated the ubiquitous presence of lignin structural units^{6,7,22}. In particular, humic substances from the Amazon River system produce high yields of lignin oxidation products characterized by low syringyl/vanillyl (S/V) and cinnamyl/vanillyl (C/V) weight ratios and relatively high (Ad/Al)_v values^{6,7}. Because the Amazon is by far the world's largest river system and contributes approximately 20% of the total freshwater input to the ocean², it serves as a good model for the expected terrestrial component within the marine DOM pool.

We chose to study water from the Pacific Ocean which is at the most extreme end of the global ocean circulation pattern and least affected by direct river input. Both surface (depth 5 m) and sub-thermocline (150 m) water samples (400 l) were collected from two stations in the eastern equatorial Pacific Ocean

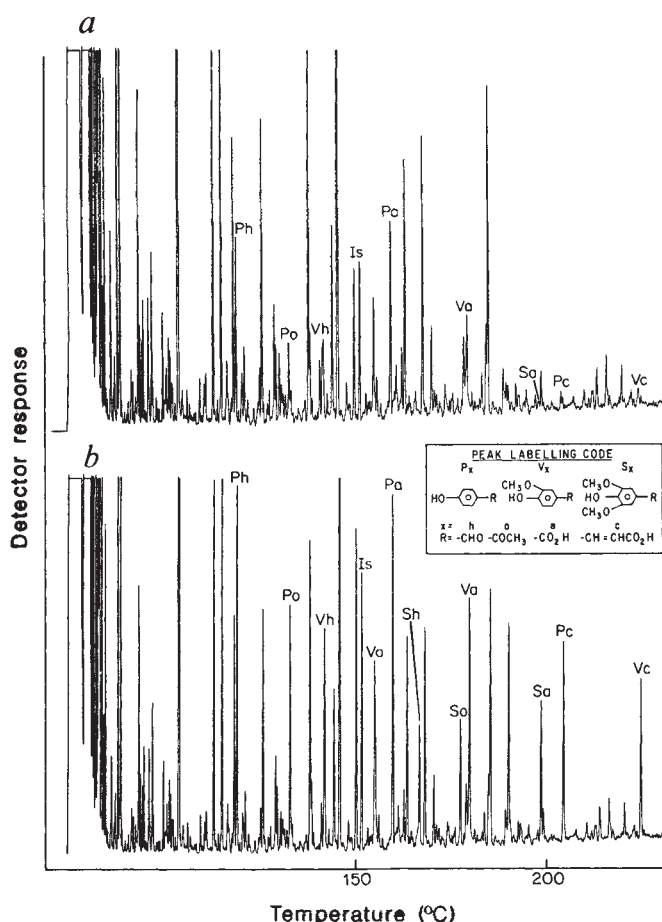


Fig. 2 Gas chromatographic traces of the cupric oxide oxidation products of the humic substance isolates from station EEP-1 (Fig. 1), sub-thermoclinic water: *a*, analysed directly as TMS derivatives; *b*, coinjected with a standard mixture of TMS-phenols. Peak identities (named as unsilylated forms): Ph, *p*-hydroxybenzaldehyde; Po, *p*-hydroxyacetophenone; Vh, vanillin; Is, ethylvanillin (GC internal standard); Vo, acetovanillon; Pa, *p*-hydroxybenzoic acid; Sh, syringaldehyde; So, acetosyringone; Va, vanillic acid; Sa, syringic acid; Pc, *p*-coumaric acid; Vc, ferulic acid. For chromatographic method, see ref. 22.

(EEP-1 and EEP-2, Fig. 1). In addition, a sample from coastal inland waters (SJF, Fig. 1; Salinity $S = 33.3\%$) was analysed. We chose to use XAD-2 (rather than XAD-8) resin to isolate and concentrate humic substances from seawater because XAD-2 gave comparable DOM recoveries in the coastal sample and produced no phenolic contamination peaks in the gas chromatographic trace of its procedural blank. The isolation procedure was carried out with 3-l resin columns as described by Thurman and Malcolm²⁴, with the exceptions that we used in-line glass-fibre filters (0.45 μm) in addition to the primary silver filters, and we required two recycling steps to concentrate the bulk humic isolate, which was not subsequently fractionated. By this method, 33.6–95.8 mg of humic material was isolated, corresponding to ~5–15% of the total dissolved organic carbon (DOC), assuming a DOC concentration of 0.8 mg l⁻¹. A procedural blank of 0.04 mg organic carbon was obtained for the entire isolation scheme, which corresponds to ~0.2% of the total organic carbon in the smallest humic isolate sample. Elemental compositions were determined in duplicate using a Carlo Erba Model 1106 CHN analyser. Stable carbon isotope compositions were determined in duplicate by Coastal Science Laboratories and are reported here in $\delta^{13}\text{C}$ notation, relative to PDB²⁵. Cupric oxide lignin oxidation products were quantified as trimethylsilyl (TMS) derivatives by gas capillary

chromatography²⁶ only after identities were verified against commercial standards by coinjection of TMS and methoxime derivatives²⁷, and by gas chromatography-mass spectrometry (GC-MS).

The average percentage of organic carbon in the humic isolates from the equatorial Pacific samples ($50 \pm 0.8\%$) agrees well with values previously reported for the open ocean¹⁵ and the Amazon River^{6,7}, and reflects a low ash content. Atomic C/N ratios of the Pacific coastal and open-ocean humic isolates (32.4–37.2) are intermediate between reported ratios for open ocean (8.63; ref. 16) and Amazon (44.6 ± 5.2 ; refs 6, 7) isolates, but are significantly higher than the previously reported oceanic value. The mean H/C ratio for the ocean samples (1.58 ± 0.07) is similar to previously determined values for seawater humics (1.57 and 1.61; refs 18, 28), confirming the aliphatic nature of this material¹⁷. The somewhat lower H/C ratio of the coastal sample (1.27) is still higher than values reported for lake (1.2; ref. 14) and Amazon River (0.88–1.1; refs 6, 7) waters. All open-ocean H/C values are much higher than that of the polystyrene XAD-2 resin (1.09), indicating minimal sample contamination by the resin. The $\delta^{13}\text{C}$ values for the marine humic substances (-21.8 to -22.2%) are within the range of values reported for humic isolates from coastal and open-ocean waters (-23.7% and -22.8% , respectively)¹⁶, and are also consistent with reported values for bulk DOC from the south (-19.9 to -22.2%)²⁹ and north-east (-21.2 to -24.2%)³⁰ Pacific Ocean. These $\delta^{13}\text{C}$ values for marine DOM are all more positive than reported values for humic substances from streams and rivers (-25.9 to -26.6%)¹⁴ and for bulk XAD-2 resin (-28.0%), confirming the predominance of uncontaminated marine-derived material in the seawater humic isolates.

Gas chromatographic traces of the cupric oxide oxidation products for all Pacific samples (coastal and open-ocean) are remarkably similar and are dominated by non-lignin-derived compounds (Fig. 2). These traces have many peaks in common with the cupric oxide oxidation product mixtures of Amazon humic isolates^{6,7}, although some additional and as yet unidentified major peaks appear in the marine samples. For the coastal marine humic sample, all characteristic lignin oxidation products (aldehydic, ketonic and acidic forms of vanillyl and syringyl phenols, plus two acidic cinnamyl phenols) can be clearly identified by their mass spectra. In the open-ocean samples, neither syringaldehyde nor acetosyringone can be convincingly discriminated, and although the presence of acetovanillon is verified by GC-MS, it cannot be sufficiently separated from other compounds to be quantified in the gas chromatographic traces (Fig. 2).

In all five marine humic isolate samples, vanillyl phenols are the dominant lignin products, contributing an average $58 \pm 5\%$ to the total definable lignin yield. The S/V and C/V weight ratios of the coastal marine sample (0.21 and 0.10, respectively)

Table 1 Elemental and lignin compositions of humic substances isolated from coastal and open-ocean waters

Station	Depth (m)	%OC	C/N	H/C	Λ	(Ad/Al) _v	$\delta^{13}\text{C}$ (‰)
SJF	100	45.0	34.5	1.27	0.19	1.3	-22.2
EEP-1	5	49.9	32.4	1.65	0.16	1.3	-22.2
EEP-1	150	50.0	37.2	1.55	0.11	2.1	-21.8
EEP-2	5	51.1	35.7	1.57	0.12	1.1	-22.0
EEP-2	150	51.5	35.8	1.56	0.11	1.5	-22.2

SJF, Strait of Juan de Fuca sample, $48^\circ 13.2' \text{N}$, $123^\circ 21.1' \text{W}$; EEP-1, eastern equatorial Pacific station 1, $2^\circ 06.1' \text{N}$, $110^\circ 09.9' \text{W}$; EEP-2, eastern equatorial Pacific station 2, $5^\circ 00.9' \text{S}$, $124^\circ 58.4' \text{W}$; %OC, weight per cent organic carbon of freeze-dried humic isolates (uncorrected for ash); C/N and H/C, atomic ratios of freeze-dried isolates; Λ , sum of the eight lignin oxidation products in mg per 100 mg organic carbon; (Ad/Al)_v, weight ratio of vanillic acid to vanillin; $\delta^{13}\text{C}$ of freeze-dried humic isolates is given relative to PDB standard.

are very close to the values for the Amazon River system ($S/V = 0.22-0.47$; $C/V = 0.01-0.10$)^{6,7}. The $(Ad/Al)_V$ ratios of all the marine humic isolates (1.1-2.1) also lie within the range for dissolved humic substances from the Amazon River system (1.1-2.2; refs 6, 7) but are, on average, higher than values reported for dissolved humic isolates from temperate freshwaters (0.79-0.98; ref. 22). The similar high $(Ad/Al)_V$ ratios for the marine and Amazon humic isolates reveal extensive oxidative biodegradation of the component lignin structural units. In all five marine samples, *p*-hydroxyl phenols are abundant relative to vanillyl compounds and *p*-hydroxyacetophenone is depleted, indicating a predominantly non-lignin source^{6,7}.

Because the relative yields of lignin oxidation products from the oceanic humic isolates are low, their quantification by means of gas chromatographic integration is crude. However, we obtain yields of total lignin phenols normalized to 100 mg of organic carbon (Λ) for open ocean humic materials that are uniformly near 0.14 (Table 1). These Λ -values are low compared with those reported for lake (2.5-3.4; ref. 22) and Amazon (0.90-2.1; refs 6, 7) humic isolates, and correspond to only 0.2% of the bulk humic carbon present as lignin structural units. Thus, the concentration of chemically recognizable lignin present in these marine humic substances is too low to account for a significant fraction of the aromatic carbon indicated by previous spectroscopic studies^{17,18}. Assuming that the lignin content of riverine humic substances is unaltered at sea, the ratio $\Lambda_{\text{ocean}}/\Lambda_{\text{Amazon}}$ (0.14/1.2; refs 6, 7) provides a crude estimate that approximately 10% of open-ocean humic material is terrestrially derived. At the minimum (assuming all terrestrially derived material is confined to this 10% humic-fraction of DOC) this corresponds to roughly 1% of the bulk DOC, with a probably upper limit of 10%, if the resin-isolated material is typical of DOC in general.

With regard to the global carbon cycle, further data is necessary to better quantify the terrestrial contribution to oceanic DOM. Additional investigations of lignin-derived phenols from various ocean-water masses and geographical locations may provide compositional trends related to locations, patterns and rates of lignin removal as well as information applicable to oceanic mixing processes.

We thank J. Ertel for advice and consultation; T. Bates and P. Murphy (PMEL/NOAA) for providing cruise equipment and GC-MS use; E. M. Thurman for instruction on XAD use; B. Clark (NWFC/NOAA) for logistical support; Rainier Brewing Company for donating kegs which were crucial to our work; and the officers and crew of the NOAA ship *Discoverer* for cooperation and help. This research was supported by NSF grants OCE-821-9294 and OCE-842-1023. This is contribution no. 1613 from the School of Oceanography, University of Washington.

Received 25 September 1985; accepted 12 February 1986.

- Likens, G. E., Bormann, F. H. & Johnson, N. M. in *Some Perspectives of the Major Biogeochemical Cycles—Scope 17* (ed. Likens, G. E.) 93-112 (Wiley, New York, 1981).
- Meybeck, M. *Am. J. Sci.* **282**, 401-450 (1982).
- Hedges, J. I. & Parker, P. L. *Geochim. cosmochim. Acta* **40**, 1019-1029 (1976).
- Hedges, J. I. & Mann, D. C. *Geochim. cosmochim. Acta* **43**, 1809-1818 (1979).
- Ertel, J. R. & Hedges, J. I. *Geochim. cosmochim. Acta* **48**, 2065-2074 (1984).
- Ertel, J. R. thesis, Univ. Washington (1985).
- Ertel, J. R., Hedges, J. I., Devol, A. H., Richey, J. E. & Ribeiro, N. *Limnol. Oceanogr.* (in the press).
- Williams, P. M., Oeschger, H. & Kinney, P. *Nature* **224**, 256-259 (1969).
- Mantoura, R. F. C. & Woodward, E. M. S. *Geochim. cosmochim. Acta* **47**, 1293-1309 (1983).
- Pempkowiak, J. & Pocklington, R. in *Aquatic and Terrestrial Humic Materials* (eds Christman, R. F. & Gjessing, E. T.) 371-386 (Ann Arbor Science, Ann Arbor, 1983).
- Lee, C. & Bada, J. L. *Earth planet. Sci. Lett.* **26**, 61-68 (1975).
- Gagosian, R. B. *Geochim. cosmochim. Acta* **39**, 1443-1454 (1975).
- Mopper, K. *Mar. Chem.* **5**, 585-603 (1977).
- Thurman, E. M. in *Organic Geochemistry of Natural Waters* (eds Kramer, C. J. M. & Duinker, J. C.) (Martinus Nijhoff/Dr W. Junk, Dordrecht, 1985).
- Stuermer, D. H. & Harvey, G. R. *Nature* **250**, 480-481 (1976).
- Gagosian, R. B. & Stuermer, D. H. *Mar. Chem.* **5**, 605-632 (1977).
- Harvey, G. R., Boran, D. A., Chesal, L. A. & Tokar, J. M. *Mar. Chem.* **12**, 119-132 (1983).
- Stuermer, D. H. & Payne, J. R. *Geochim. cosmochim. Acta* **40**, 1109-1114 (1976).
- Wilson, M. A., Philp, R. P., Gillam, A. H., Gilbert, T. D. & Tate, K. R. *Geochim. cosmochim. Acta* **47**, 497-502 (1983).
- Sarkanen, K. V. & Ludwig, C. H. (eds) *Lignins* (Wiley, New York, 1971).

- Crawford, R. L. (ed.) *Lignin Biodegradation and Transformation* (Wiley, New York, 1981).
- Ertel, J. R., Hedges, J. I. & Perdue, E. M. *Science* **223**, 485-487 (1984).
- Hedges, J. I., Ertel, J. R. & Leopold, E. B. *Geochim. cosmochim. Acta* **46**, 1869-1877 (1982).
- Thurman, E. M. & Malcolm, R. L. *Envir. Sci. Technol.* **15**, 463-466 (1981).
- Degens, E. T. in *Organic Geochemistry* (eds Eglinton, G. & Murphy, M. T. J.) 304-329 (Springer, New York, 1969).
- Hedges, J. I. & Ertel, J. R. *Analyt. Chem.* **54**, 174-178 (1982).
- Horning, M. G., Moss, A. M. & Horning, E. C. *Analyt. Biochem.* **22**, 284 (1968).
- Gillam, A. H. & Wilson, M. A. *Org. Geochem.* **8**, 15-25 (1985).
- Eadie, B. J., Jeffrey, L. M. & Sackett, W. M. *Geochim. cosmochim. Acta* **42**, 1265-1269 (1978).
- Williams, P. M. & Gordan, L. *Deep Sea Res.* **17**, 19-27 (1970).

Fluorescently labelled Na^+ channels are localized and immobilized to synapses of innervated muscle fibres

Kimón J. Angelides*

Department of Neuroscience and Center for Neurobiological Studies, University of Florida College of Medicine, Gainesville, Florida 32610, USA

Segregation of voltage-dependent sodium channels to the hillock of motoneurons and nodes of Ranvier in myelinated axons is crucial for conduction of the nerve impulse^{1,2}. Much less is known, however, about the distribution of voltage-dependent Na^+ channels on muscle fibres. Recently, Beam *et al.*³ have shown that Na^+ channels are concentrated near the neuromuscular junction. To determine the topography and mechanisms governing the distribution of voltage-dependent Na^+ channels on muscle, micro-fluorimetry and fluorescence photobleach recovery (FPR) have now been used to measure the density and lateral mobility of fluorescently labelled Na^+ channels on uninnervated and innervated muscle fibres. On uninnervated myotubes, Na^+ channels are diffusely distributed and freely mobile, whereas after innervation the channels concentrate at neuronal contact sites. These channels are immobile and co-localize with acetylcholine receptors (AChRs). At extrajunctional regions the Na^+ channel density is lower and the channels more mobile. The results suggest that the nerve induces Na^+ channels to redistribute, immobilize and co-localize with AChRs at sites of neuronal contact.

Sodium channels were fluorescently labelled by exploiting the high affinity of fluorescent derivatives of toxin V from *Leiurus quinquestratus* (Lq V), toxin II from *Centruroides suffusus* (Css II) and tetrodotoxin (TTX)⁴⁻⁶. Topographical mapping and lateral diffusion coefficients (D_L) of labelled Na^+ channels on aneurally cultured embryonic chick myotubes, chick spinal cord neurone-muscle co-cultures and mouse diaphragm muscle were measured using the spot photobleaching technique⁷.

Figure 1 shows the fluorescence labelling of voltage-dependent Na^+ channels by 5-(6-)carboxytetramethylrhodamine-Lys-Lq V (TmRhd-Lq V) on aneurally cultured chick myotubes and myotubes co-cultured with spinal cord neurones. Fluorescence staining of non-innervated myotubes was faint and diffusely distributed over the cell surface (Fig. 1a) whereas in the nerve-muscle co-cultures (Fig. 1b), more localized and patchy fluorescence was observed at regions which appeared morphologically identifiable as neuronal contact regions and which coincided with fluorescein 5-isothiocyanate (FITC)- α -bungarotoxin fluorescence^{8,9}. Comparison of Na^+ channel and AChR distribution showed that the two channels occupy contiguous regions (lower panel to Fig. 1b). In most fibres, Na^+ channel and AChR staining appeared as single fluorescent patches that co-localized at neuronal contact sites, although Na^+ channel fluorescence was characteristically less intense than the AChR fluorescence. AChR patches were occasionally found in regions not associated with the neurite. Patches of fluorescently labelled Na^+ channels did not co-localize with these neurite-free AChR clusters.

* Present address: Dept. of Physiology and Molecular Biophysics, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA.

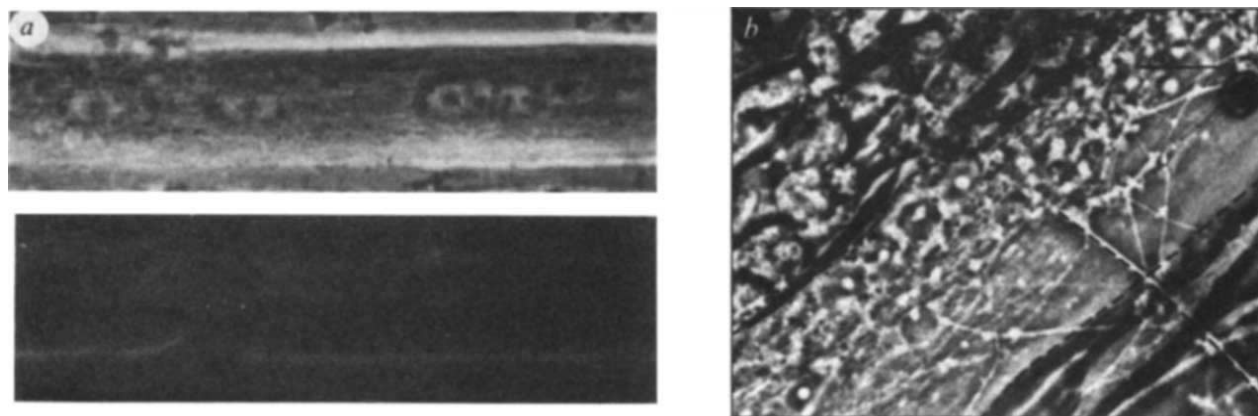


Fig. 1 Fluorescence labelling of Na^+ channels by TmRhd-Lqq V in 11-day-old chick myotubes (a) and chick myotubes innervated by spinal cord neurones (b). TmRhd-Lqq V binds reversibly to these cells with a K_d of 16 nM. The half-time for dissociation is 65 min at 22 °C. The upper micrographs are phase-contrast images of portions of the TmRhd-Lqq V-labelled myotubes and spinal cord neurone-muscle cell co-cultures. The lower micrographs are fluorescence labelling of Na^+ channels on 11-day-old chick myotubes. In b AChRs were visualized on the same fibre after labelling with 10^{-7} M FITC- α -bungarotoxin followed by washing. The bottom panel in b is a colour overlay of the AChR and Na^+ channel staining patterns on the same fibre. Scale bar, 10 μm . **Methods.** The cells were incubated with 40 nM TmRhd-Lqq V for 30 min at 22 °C in a buffer composed of 130 mM choline chloride, 50 mM HEPES adjusted to pH 7.4 with Tris base, 5.5 mM glucose, 0.8 mM MgSO_4 , 5.4 mM KCl, to which 2 μM TTX was added to block any ion movement through the Na^+ channels that would lead to membrane depolarization and dissociation of TmRhd-Lqq V (ref. 4). The cultures were rapidly washed five times with ice-cold buffer before photography. Nonspecific binding of ^{125}I -Lqq V to these cells was 15% after depolarization by replacing the external medium with 150 mM external K^+ or by competitive displacement with 200 nM unlabelled Lqq V. The photographs were taken on Ektachrome film pushed to 1,200 ASA using a Nikon UFX automatic exposure attachment through a $\times 40$ water immersion, 0.75 numerical aperture objective or a $\times 100$ glycerin immersion, 1.4 numerical aperture objective.

To obtain a detailed profile of Na^+ channel distribution, fluorescently labelled Na^+ channels were mapped with a laser microbeam at various distances from the site of nerve contact using the AChR cluster identified by FITC- α -bungarotoxin as the origin. Figure 2 shows the distribution of these channels. Na^+ channel density in such cultures is much higher at the site of innervation. Within 20 μm of this region the fluorescence intensity doubles and at 200 μm it falls to $\sim 15\%$ of that of the synaptic region. Examination of AChR distribution shows that within the same region the drop in FITC- α -bungarotoxin fluorescence is three- to fivefold greater than the drop in TmRhd-Lqq V fluorescence. When FITC-succinyl concanavalin A (FITC-sCon-A) or tetramethylrhodamine phosphatidylethanolamine (TmRhd-PE) receptors were examined, no pattern in the fluorescence intensities was observed, indicating that the steep gradient observed for Na^+ channels is not due to local invaginations of the membrane at the nerve-muscle junction. Furthermore, this suggests that, compared with the Na^+ channel protein, the general classes of phospholipids and glycoproteins are more homogeneously distributed on the muscle cell surface. The absence of patches or a steep concentration gradient of FITC-sCon-A receptors co-localizing with Na^+ channels or AChRs shows that the innervated sarcolemma forms and maintains a distinct distribution of different membrane proteins. Beam *et al.*³ and Roberts and Almers¹⁰ have shown that on adult innervated skeletal muscle fibres, Na^+ channels congregate at the neuromuscular junction. Using the loose patch-clamp technique, they observed that immediately adjacent to the endplate the Na^+ current density, which they associated with voltage-dependent Na^+ channels, was 5–10-fold higher than at regions away from the endplate. Although the increased Na^+ conductance at the synapse could arise from channels possessing a high single-channel conductance, this now seems unlikely because our independent measurements by fluorescence methods using probes specific for the voltage-dependent Na^+ channel confirm that Na^+ channels are concentrated at synapses.

To gain insight into the mechanisms that organize Na^+ channels at the synapse and the elements maintaining this distribution, the lateral mobility of fluorescently labelled Na^+ channels was measured by FPR at several locations on the muscle cell. Figure 3 shows representative fluorescence recovery curves obtained from uninnervated myotubes (Fig. 3a) and myotubes

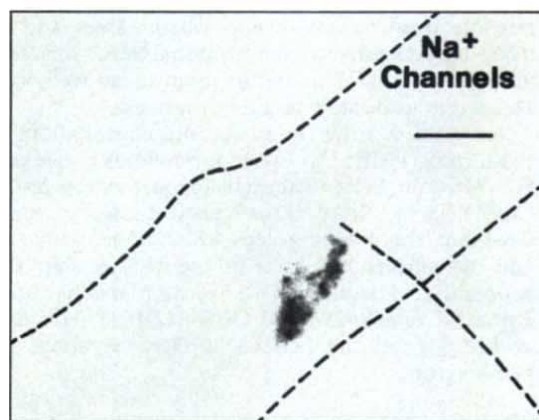
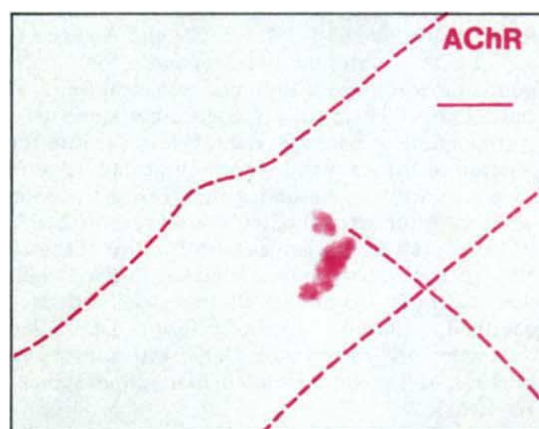


Table 1 Lateral diffusion coefficient (D_L) and mobile fraction (f) of sodium channels, FITC-sCon-A receptors and TmRhd-PE on muscle fibres

Cell type and distance from AChR cluster (μm)	Sodium channel*			FITC-sCon-A			TmRhd-PE		
	$\langle D_L \rangle \pm D_L$ $\times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$	$\langle f \rangle \pm f$	N	$\langle D_L \rangle \pm D_L$ $\times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$	$\langle f \rangle \pm f$	N	$\langle D_L \rangle \pm D_L$ $\times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$	$\langle f \rangle \pm f$	N
Myotube	8.9 ± 2.5	1.04 ± 0.14	14	1.1 ± 0.20	0.57 ± 0.13	14	0.81 ± 0.15	0.93 ± 0.11	14
Spinal/muscle									
0	IM	0.02 ± 0.02	7	2.4 ± 0.37	0.11 ± 0.08	4	0.50 ± 0.08	0.91 ± 0.13	4
14	IM	0.073 ± 0.05	7	1.8 ± 0.11	0.12 ± 0.06	4	0.70 ± 0.08	0.96 ± 0.08	4
	(NBD-TTX:IM)	0.05 ± 0.05	2						
20	0.22 ± 0.09	0.11 ± 0.06	9	2.4 ± 0.26	0.08 ± 0.04	4	0.68 ± 0.09	0.67 ± 0.23	4
54	0.39 ± 0.08	0.34 ± 0.1	7	1.6 ± 0.17	0.27 ± 0.11	4	0.72 ± 0.11	0.73 ± 0.17	4
112	1.71 ± 0.53	0.39 ± 0.08	7	1.3 ± 0.27	0.37 ± 0.10	4	0.91 ± 0.11	0.83 ± 0.15	4
	(C55: 1.31 ± 0.43)	0.42 ± 0.10	3						
240	2.3 ± 1.2	0.42 ± 0.06	11	1.89 ± 0.39	0.39 ± 0.14	4	0.86 ± 0.09	0.76 ± 0.09	4
Adult muscle fibres									
Endplate	0.039 ± 0.019	IM	6	IM	0.057 ± 0.05	6	0.82 ± 0.07	0.74 ± 0.04	6
200 μm	2.81 ± 0.42	0.55 ± 0.24	6	1.64 ± 0.11	0.41 ± 0.17	8	0.91 ± 0.09	0.75 ± 0.07	6

$\langle D_L \rangle$, Lateral diffusion coefficient $\pm D_L$, s.d.; $\langle f \rangle$, mobile fraction $\pm f$, standard deviation. IM, immobile on FPR timescale, $D \leq 10^{-12} \text{ cm}^2 \text{ s}^{-1}$. N, number of measurements.

* Measured with TmRhd-Lqq V, unless otherwise noted using NBD-TTX or TmRhd-C55 II.

innervated by spinal cord neurones (Fig. 3b, c). The recovery curves show that on uninnervated myotubes all Na^+ channels are relatively unrestricted in their lateral mobility, with diffusion coefficients of $8.9 \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$ (Fig. 3a) whereas Na^+ channels clustered at synapses are immobilized, with diffusion coefficients of $\leq 10^{-12} \text{ cm}^2 \text{ s}^{-1}$ (Fig. 3b). However, not all Na^+ channels are immobile. For example, at 240 μm from the site of neuronal contact, which is a region of lower channel density, 40% of the Na^+ channels are mobile, with lateral diffusion coefficients of $2.3 \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$ (Table 1).

A general trend emerged from these experiments. Those channels located at and close to neuronal contact sites are immobile, while the rate of lateral motion and fraction of Na^+ channels able to move increase with increasing distance from the synapse. Similar observations were made on adult mouse diaphragm muscle, so these observations do not seem to be restricted to the culture system. The use of three Na^+ channel probes which bind to the Na^+ channel protein at independent and topologically remote sites^{4,6}, and the demonstration that the diffusion coefficients are independent of the probes themselves, eliminate any biasing or artefactual contribution from the probes. As an additional control, the fluorescent phospholipid derivative TmRhd-PE diffuses freely in all membrane regions with diffusion coefficients of $7\text{--}10 \times 10^{-9} \text{ cm}^2 \text{ s}^{-1}$ (Table 1). *t*-tests of the data show that there is no statistical difference between these regions. Therefore, inhomogeneities in membrane lipid viscosity cannot account for the different channel diffusion rates observed at locations on and away from the neuronal contact site.

Although the diffusion coefficient for FITC-sCon-A receptors does not change over the muscle cell surface, the mobile fraction increases somewhat in parallel with the Na^+ channel distribution (Table 1). The immobility of many FITC-sCon-A receptors at the synapse may be due in part to the high density of AChRs and Na^+ channels, both of which are glycoproteins and sCon-A receptors¹¹⁻¹³. However, this region contains a small but important group of sCon-A receptors, besides AChRs or Na^+ channels, which are mobile. This indicates that not all proteins in this region are immobile and that Na^+ channels experience still stronger restrictions in lateral mobility.

The results using these probes demonstrate that the heterogeneous distribution and differential mobility of Na^+ channels are not a characteristic of all sarcolemmal proteins, the result of differences in membrane fluidity created by partitioning of particular phospholipids¹⁴, or domains created by the aggregation and clustering of general classes of intrinsic membrane proteins. Our FPR results using nerve-muscle co-cultures and Na^+ channel-specific fluorescent probes differ slightly from those of Stuhmer and Almers¹⁵, for they found, using ultraviolet photobleaching and electrophysiological techniques, that on frog muscle all Na^+ channels were immobile.

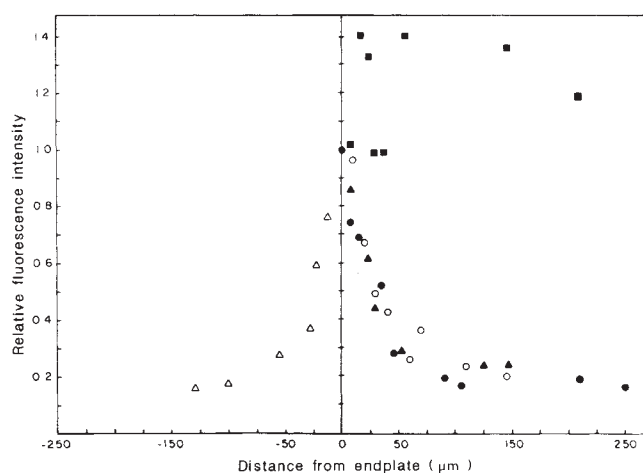


Fig. 2 Profile of fluorescently labelled Na^+ channel (●, ○, ▲, △) and FITC-sCon-A receptor (■) distribution on myotubes innervated by spinal cord neurones *in vitro*. Typically, a nerve process from a large multipolar neurone was visualized as it passed on top of, adjacent to, or terminated on the muscle fibre by phase or interference contrast illumination. Regions which appeared morphologically identifiable as neuronal contact sites were then examined for the localization of AChRs with FITC- α -bungarotoxin^{8,9}. The monitoring beam of an argon ion laser was used to excite a small spot (1.7 or 3.6 μm diameter) on the cell membrane at various regions of the muscle and the fluorescence was accumulated at a photon counting interval of 40 ms. The beam diameter and a reticulated eyepiece were used to record the position of the microbeam relative to the nerve-muscle junction. Within each experiment no change in neutral density filters or in the optical arrangement was made. The neurite-associated ACh receptor patch at the junctional region was defined as the origin (zero distance). Photon counts were normalized to the value obtained at the neurite-associated ACh receptor patch at the junctional region, which was 1,254 counts per 40 ms for specific TmRhd-Lqq V binding and 990 counts per 40 ms for FITC-sCon-A. The absolute photon count levels, however, cannot be compared between FITC-sCon-A and TmRhd-Lqq V because a higher neutral density was required to reduce FITC-sCon-A pre-bleaching. Four representative muscle fibres are shown by the open and closed circles and triangles. The nonspecific binding was measured by displacing TmRhd-Lqq V with 200 nM unlabelled Lqq V and re-measuring the fluorescence intensities under identical counting conditions. The nonspecific binding (15%) was subtracted from the total binding to yield the specific binding to Na^+ channels⁴. The absence of a steep gradient in the fluorescence profile of FITC-sCon-A receptors demonstrates that the high fluorescence intensity observed for Na^+ channels is not due to localized foldings of the membrane which could enhance the fluorescence intensity by increasing the effective surface area.

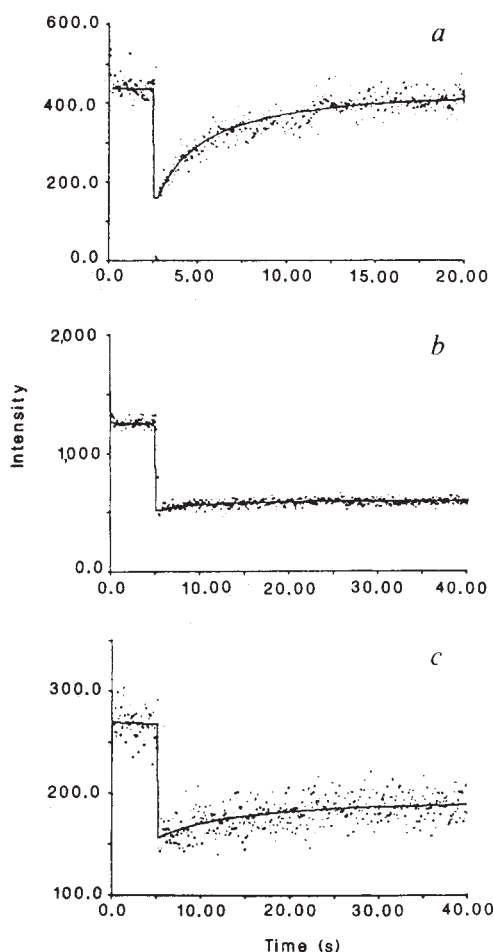


Fig. 3 Typical fluorescence recovery curves demonstrating Na⁺ channel mobility at myotubes (a), immobility at the neuromuscular junction (b) and slow mobility 240 μm from the neuromuscular junction (c) in spinal cord neurone-muscle co-cultures. Curves b and c were obtained on the same muscle fibre. D_L and f were determined by curve-fitting procedures⁷. Values for D_L (cm² s⁻¹) were determined using the expression $w^2/4T_D$ in which w is the e⁻² radius of the beam profile and T_D is the half-time for recovery. Incomplete fluorescence recovery was interpreted as an indication that a fraction of labelled channels is immobile on the timescale of the experiment; $D_L \leq 3 \times 10^{-12}$ cm² s⁻¹ was considered to indicate immobility. The radius of the bleaching spot for the $\times 100$ oil immersion objective used in this figure was 0.85 μm. Analysis of the curve in a yields a diffusion coefficient of 8.9×10^{-10} cm² s⁻¹ and a recoverable fraction of 1.00. In b, $D_L \leq 10^{-12}$ cm² s⁻¹, and in c, $D_L = 2.3 \times 10^{-10}$ cm² s⁻¹ and $f = 0.37$.

In addition to the possibility of species differences, differences in the maturity of the tissue may contribute to the results observed. The nerve-muscle cultures may be at a stage early in development in which channels at extrajunctional regions still have some lateral mobility.

The sequelae of Na⁺ channel accumulation at newly formed synapses is reminiscent of AChR cluster formation, where diffusely distributed AChRs cluster and become immobilized to the neuromuscular junction¹⁶. However, in contrast to AChRs, there do not appear to be regions of preexisting, dense, immobile Na⁺ channel patches on uninnervated myotubes. Rather, all Na⁺ channels are diffusely distributed and very mobile before innervation. The factors that specifically control the reorganization and subsequent immobilization of Na⁺ channels are unknown. Apparently AChR clustering *per se* does not induce Na⁺ channel clustering. Further, the specific recruitment of Na⁺ channels and the conspicuous absence of Na⁺ channels co-localizing with all AChR hotspots suggest that the factors

governing Na⁺ channel cluster formation show some degree of temporal and molecular specificity. However, it is not clear whether the redistributed Na⁺ channels originate from areas of diffuse Na⁺ channel distribution on the uninnervated myotube or from an endogenous intracellular pool of newly synthesized Na⁺ channels which preferentially appear at the synapse during the time of neuronal contact.

The localization and immobilization of Na⁺ channels to the synapse is unexpected, for in muscle the action potential is propagated throughout the length of the fibre. The functional role of the high Na⁺ channel density at the synapse is unclear but it has been proposed to raise the safety factor for impulse transmission³. Further experiments are required to elucidate the mechanisms by which the nerve induces the redistribution of Na⁺ channels and the specific intra- and extracellular elements that maintain the elevated Na⁺ channel density at the synapse.

I thank Dr Elliot Elson and David Loftus for use of the FPR equipment and for helpful discussions, the laboratory of Dr G. Fischbach, Washington University, for supplying some of the co-cultures, and Dr John Caldwell for careful reading of the manuscript and for invaluable comments. This work was supported by grants from the NIH (NS 18268) and the Muscular Dystrophy Association.

Received 7 October 1985; accepted 10 March 1986.

1. Coombs, J. S., Eccles, J. C. & Fatt, P. J. *Physiol., Lond.* **130**, 291-325 (1955).
2. Waxman, S. G. & Ritchie, J. M. *Science* **228**, 1502-1507 (1985).
3. Beam, K. G., Caldwell, J. H. & Campbell, J. T. *Nature* **313**, 588-590 (1985).
4. Angelides, K. J. & Nutter, T. J. *J. biol. Chem.* **258**, 11948-11957 (1983).
5. Darbon, H. & Angelides, K. J. *J. biol. Chem.* **259**, 6074-6084 (1984).
6. Angelides, K. J. *Biochemistry* **20**, 4107-4118 (1981).
7. Axelrod, D. A., Koppel, D. E., Schlessinger, J., Elson, E. L. & Webb, W. W. *Biophys. J.* **16**, 1055-1069 (1976).
8. Ravdin, P. & Axelrod, D. A. *Analyt. Biochem.* **58**, 585-592 (1977).
9. Role, L. W., Matossian, V. R., O'Brien, R. J. & Fischbach, G.-D. *J. Neurosci.* **5**, 2197-2204 (1985).
10. Roberts, W. M. & Almers, W. *Biophys. J.* **47**, abstr. 189 (1985).
11. Brookes, J. P. & Hall, Z. W. *Biochemistry* **14**, 2100-2106 (1975).
12. Cohen, S. & Barchi, R. L. *Biochim. biophys. Acta* **645**, 253-261 (1981).
13. Elmer, L. W., O'Brien, B., Nutter, T. J. & Angelides, K. J. *Biochemistry* **24**, 8128-8137 (1985).
14. Klausner, R. D., Kleinfeld, A. M., Hoover, R. L. & Karnovsky, M. J. *J. biol. Chem.* **255**, 1286-1295 (1980).
15. Stuhmer, W. & Almers, W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 946-950 (1983).
16. Stya, M. & Axelrod, D. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 449-453 (1983).

Single cyclic GMP-activated channel activity in excised patches of rod outer segment membrane

L. W. Haynes, A. R. Kay & K.-W. Yau

Department of Physiology and Biophysics, University of Texas Medical Branch, Galveston, Texas 77550, USA

The plasma membrane of retinal rod outer segments contains a cyclic GMP-activated conductance¹⁻⁶ which appears to be the light-sensitive conductance involved in phototransduction⁷. Recently, it has been found that this conductance is partially blocked by Mg²⁺ (refs 3, 8, 9) and Ca²⁺ (refs 3, 4, 8-11) at physiological concentrations, thus possibly accounting for the absence of observable single-channel activity in excised membrane patches^{1,3,8} and for the unusually small apparent unit conductance deduced from noise measurements on intact cells^{4,10,12,13}. We now report that, as expected from this idea, single cGMP-activated channel activity can be detected from an excised rod membrane patch in the absence of divalent cations. The most prominent unitary current had a mean conductance of ~25 pS. Both individual channel openings (mean open time ~1 ms) and short bursts of openings (mean burst duration of about a few milliseconds) were observed. In addition, there were smaller events which probably

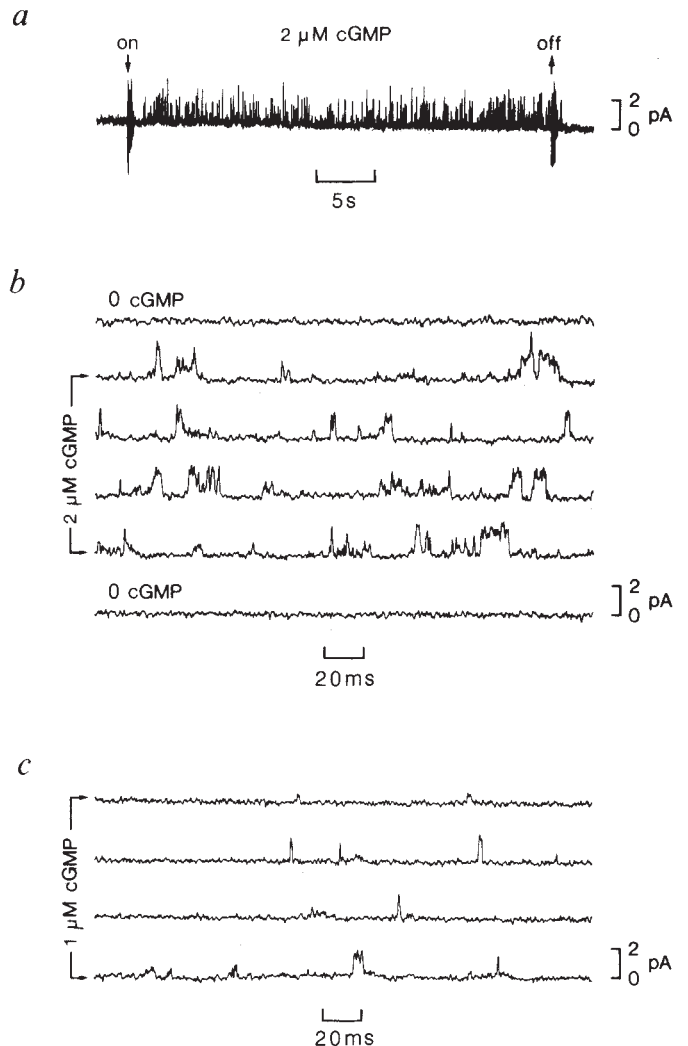
Fig. 1 Single cGMP-activated channels recorded from an inside-out membrane patch excised from a toad rod outer segment (ROS). Isotonic NaCl solution (118 mM NaCl, 0.1 mM Na-EGTA, 0.1 mM Na-EDTA, 5 mM Na-HEPES, pH 7.6) was present on both sides of the membrane. The holding potential was +60 mV (that is, interior of pipette being negative). All records were low-pass-filtered at 2 kHz (8-pole Bessel). *a*, Activation of single-channel activity by 2 μ M cGMP. The 'on' and 'off' arrows indicate the times when tap movements were made to admit cGMP into the bath and to remove it. The ringing transients immediately following the on and off time markers were caused by vibrations from the tap movements. *b*, Samples of the same recording shown at a higher time resolution. Top trace, control recording before the admission of cGMP; bottom trace, control recording after removal of cGMP. The middle four traces, obtained in the presence of 2 μ M cGMP, were not continuous in time, but were selected to illustrate various features associated with openings and closings of the cGMP-activated channels (see text). *c*, Recordings from the same patch at 1 μ M cGMP.

Methods. All experiments, except for that in Fig. 4, were performed on ROS from the toad, *Bufo marinus*. The experiment in Fig. 4 was performed on a rod from the larval tiger salamander (*Ambystoma tigrinum*), simply because for a time we failed to obtain stable gigohm seals from toad rods; other control experiments have indicated that the cGMP-activated channels in the salamander rod have properties similar to those in the toad rod. A piece of retina from a dark-adapted animal was incubated in the dark for 1 h in Ringer's solution containing 1 mg ml⁻¹ hyaluronidase (Sigma type I-S), then finely chopped to yield many isolated rod outer segments⁸. The outer segments were then transferred to the experimental chamber and washed thoroughly with isotonic NaCl solution (see above) before recordings began. All recordings were made in visible light at room temperature (23–25 °C). The patch pipette, containing isotonic NaCl solution, was made of thick-walled borosilicate glass capillary and had a lumen diameter at the tip of ≤ 0.5 μ m. The exterior of the pipette tip was coated with Sylgard to reduce electrical noise³⁰. Current was measured with a List EPC-7 patch-clamp system in the voltage-clamp mode. Gigohm seals were made on the plasma membrane of an outer segment near its tip, and the patch was then excised in the inside-out configuration by a gentle tap on the micromanipulator supporting the pipette^{2,3,8}. The signals were stored on analog tape and subsequently digitized for analysis.

represented other states of the conductance. The mean current increased with the third power of cGMP concentration, suggesting that there are at least three cGMP-binding sites on the channel molecule. With 0.2 mM Mg²⁺ in the cGMP-containing solution, a flickering block of the open channel was observed; the effect of Ca²⁺ was similar. The results resolve a puzzle about the light-sensitive conductance by demonstrating that it is an aqueous pore rather than a carrier.

All experiments were done in visible light, with an isotonic NaCl solution (see Fig. 1 legend) present on both sides of the membrane. Figure 1*a* shows the current recorded from an excised, inside-out rod membrane patch held at +60 mV. With 2 μ M cGMP in the bath, there were fluctuations which resembled the openings and closings of individual ion channels. The effect of cGMP was rapidly reversible. Figure 1*b* shows sample records at a higher time resolution. The most prominent unitary currents, which occurred both singly and in short bursts, had a mean amplitude of ~ 1.5 pA (see Fig. 2*a*), corresponding to a conductance of ~ 25 pS. Apart from the large unit currents there were also small events, but these were more difficult to analyse because their amplitudes were not significantly above background noise. The relationship between the small and the large events is not clear at present, though they probably represent different states of the same conductance. In Fig. 1*c*, the same patch was exposed to a lower concentration of cGMP: the frequency of events was lower, and the bursts of openings also seemed to be briefer.

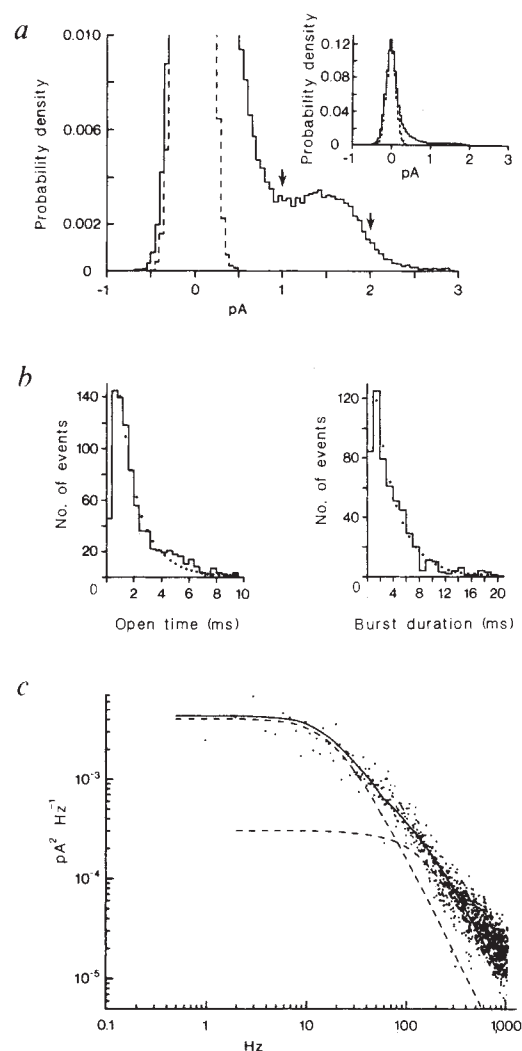
Figure 2*a* shows probability density histograms of current amplitude obtained from the recordings illustrated in Fig. 1*a*



and *b*. The solid profile is the histogram obtained in the presence of 2 μ M cGMP, while the dashed profile is the control histogram scaled to the same height. The former histogram has two identifiable peaks: one centred at zero and the other at ~ 1.5 pA. As judged from the raw records, the non-zero peak appears to correspond to the mean amplitude of the large unitary events. The distribution centred at 0 pA overlapped with the control histogram quite well at negative current values, but at positive values there was excess probability density which reflected the contribution of the small events. From the histogram, we estimated that the probability of occurrence for the small events was roughly the same as that for the large events.

We have examined the kinetics of large events selected by a window, as indicated by the arrows in Fig. 2*a*. The open times of the events (low-pass-filtered at 1 kHz; see Fig. 2 legend) showed a roughly exponential distribution, with a time constant of 1.5 ms (Fig. 2*b*, left); the true value, however, was estimated to be ~ 1.0 ms after correcting for low-pass filtering¹⁴. The burst times (Fig. 2*b*, right) were also roughly exponentially distributed, with a time constant of 3.3 ms. A similar analysis of the single-channel activity at 1 μ M cGMP (that is, Fig. 1*c*) gave a comparable channel open time (uncorrected time constant = 1.3 ms) but a shorter burst time (time constant = 2.4 ms). The significance of this concentration-dependent difference is unclear. It should be emphasized that even though there was little stacking of events at the low concentrations of cGMP used; the successive events did not necessarily reflect activity from the same channel molecule, because many channels were gen-

Fig. 2 Analysis of the single-channel records illustrated in Fig. 1a and b. The automated single-channel analysis program of Sachs *et al.*³¹, implemented by Lingle (Florida State University) for the IBM PC, was used to produce the plots shown in a and b. Before analysis, the data were first filtered at 2 kHz with an 8-pole Bessel filter and digitized at 5 kHz, then further filtered at 1 kHz with a digital Gaussian filter to reduce background noise. **a**, Probability density histogram of current amplitude (continuous profile) in the presence of 2 μM cGMP. The dashed profile indicates the control histogram obtained after the removal of cGMP; it has been scaled to the same height as the one with cGMP. To compute each histogram, the entire stretch of recording was divided into consecutive sections each 102.4 ms long. The baseline for each section (corresponding to 0 pA on the histograms) was located using an algorithm that maximizes the number of crossings through a horizontal line, and the probability density distribution (bin width 0.05 pA) was then determined. The control histogram was fitted well by a Gaussian distribution with a standard deviation of 0.12 pA. A similar control histogram obtained before the application of cGMP was also fitted by a Gaussian distribution with a standard deviation of 0.15 pA. The non-zero peak of the histogram obtained at 2 μM cGMP is situated at ~ 1.5 pA, and corresponds to the mean amplitude of the large unitary events (see text). The arrows indicate the amplitude window used for computing the time histograms for the large events (see b). Inset, the same plots at reduced scales. The probability density histogram for the recordings at 1 μM cGMP (that is, Fig. 1c) was similar to the histogram for 2 μM shown here; the non-zero peak was in the same place, but the overall probability for non-zero amplitudes was less, as expected from the relative paucity of events. **b**, Histograms of open times and burst durations for the large events. The amplitude window chosen for analysing these large events (1.0–2.0 pA) also served to minimize the inclusion of stacked events. A burst was defined as either a single-channel opening or a collection of consecutive openings separated by closed periods ≤ 2 ms. This value was deduced from the closed-time distribution, which could be fitted by the sum of two exponentials, with the shorter time constant (~ 2 ms) representing the intra-burst closures. The dots in each plot indicate a single exponential fitted to the histogram. The time constant was 1.5 ms for the open time distribution and 3.3 ms for the burst-time distribution. The fall in probability density for short times in each histogram was due to low-pass filtering. **c**, Power density spectrum for the records in Fig. 1b. The raw records in this case were low-pass-filtered at 1 kHz (8-pole Butterworth) and digitized at 2 kHz. The control spectrum obtained in the absence of cGMP has been subtracted. The dashed curves indicate Lorentzians with corner frequencies of 20 Hz and 150 Hz respectively, corresponding to time constants of 8 and 1 ms. The continuous curve is the sum of the two Lorentzians.



erally present in a patch (see below). Therefore, the closed-time distribution was of limited usefulness and will not be described here.

Figure 2c shows the power spectrum for the current records at 2 μM cGMP. It can be fitted approximately by the sum of two Lorentzians, with characteristic time constants of ~ 1 ms and 8 ms respectively. The short time constant is comparable in value to the mean open time of the large events, while the long time constant is probably related to the burst times. Interestingly, the short time constant is rather similar to values obtained by others from noise measurements on intact cells^{4,10,12,13}. It is difficult, however, to make an exact comparison between the power spectral analysis and the kinetic analysis because, among other reasons, both large and small events contributed to the power spectrum.

An analysis similar to those in Fig. 2a, b, performed on data from three other patches (all at +60 mV and 1–2 μM cGMP), gave comparable mean unit amplitudes (1.4–1.5 pA) and time constants (event open time 1.4–1.6 ms, burst duration 1.6–3.3 ms, 1-kHz filter). As a comparison, the power spectra obtained from four experiments (1–5 μM cGMP, +60 mV) gave a short time constant of 0.9–1.5 ms and a long time constant of 7–11 ms. At frequencies ≥ 1 kHz, the power spectra often showed some residual power density which could not be accounted for by just two Lorentzians. In experiments where channel openings were also studied at –60 mV (see, for example, Fig. 3c), the events were inverted in polarity and had shorter open times (see legend to Fig. 3c). Nevertheless, the latter events had amplitudes similar to those at +60 mV. Thus, the current-voltage relation

for the unitary conductance seems to be roughly linear, which agrees with our previous conclusion from macroscopic current measurements—that is, a large part of the outward rectification observed in physiological solutions is the result of a voltage-dependent block by divalent cations^{3,8,9}.

One advantage of removing divalent cations from the solutions is that a much larger cGMP-induced current is obtained^{3,8,9}; this permits the activation of current by cGMP to be studied at very low concentrations of the activator. Figure 3a shows one such experiment performed at –30 mV. As the cGMP concentration increased, the single-channel events gradually summed to give a noisy, steady current. No obvious desensitization was evident even after prolonged (>1 min) applications of cGMP. The large amplitude of the saturated current compared with the single-channel current indicated that many channels must be present in the membrane patch, a conclusion arrived at for all the experiments that we analysed. We have attempted to estimate the channel density on the membrane, but this was complicated by the finding that the saturated current amplitude showed large variations (up to almost 100-fold) in different experiments, with little correlation to the apparent patch size (as judged visually). Assuming an average patch size of 0.2 μm^2 (as estimated from the pipette tip lumen) and assuming that at saturating cGMP concentrations each channel molecule is conducting current about half the time, the high and low estimates for the channel density would be 3,000 and 30 μm^{-2} , respectively (average $\sim 1,000 \mu\text{m}^{-2}$, 18 patches). Taking a rhodopsin density on the surface membrane of 30,000 molecules μm^{-2} , the channel/rhodopsin mole ratio

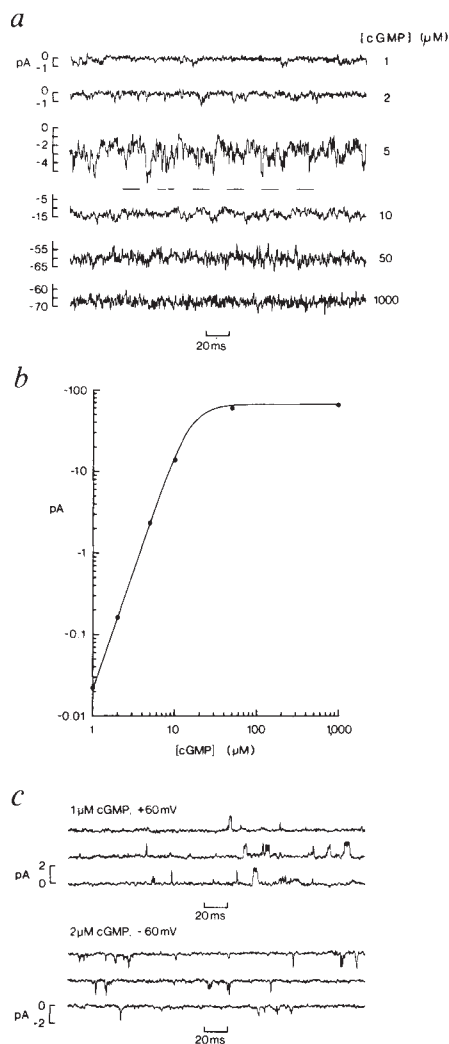


Fig. 3 *a*, Activation of current in toad rod patch at different concentrations of cGMP. Holding potential -30 mV. Records were filtered at 1 kHz. The top three and bottom three traces have different amplitude scales. Note the sustained current fluctuations even at a saturating cGMP concentration. In addition, at 50 μ M and $1,000$ μ M cGMP, there were also slow fluctuations which are not apparent in the records shown here because of the fast time scale. Total current variances (d.c. to 1 kHz) at the different cGMP concentrations were: 0.01 pA^2 (1 μ M), 0.08 pA^2 (2 μ M), 1.15 pA^2 (5 μ M), 6.33 pA^2 (10 μ M), 14.56 pA^2 (50 μ M) and 13.60 pA^2 (1 mM). *b*, log-log plot of mean current against cGMP concentration. The mean currents were computed from the records shown in *a*. Smooth curve is drawn according to the Hill equation: $j(C) = -67 \text{ pA} \times [C^3 / (C^3 + 15^3)]$, where C is the cGMP concentration. *c*, Effect of membrane potential on current activation by cGMP. Different toad rod patch from that in *a* and *b*. Only sample records are shown. Low-pass-filtered at 2 kHz. Analysis of the complete recordings indicated that the frequency of channel activation was about 2.5 times higher for 1 μ M cGMP/ $+60$ mV than for 2 μ M cGMP/ -60 mV. Open time constant (1 -kHz filtering) was 1.4 ms at $+60$ mV and 0.8 ms at -60 mV.

would be between $1:10$ and $1:1,000$ (average $\sim 1:30$). The wide range of channel density estimates suggests that there may be a local clustering of channels on the plasma membrane. Previous estimates based on noise measurements from intact cells were between 100 and 200 channels per μm^2 (refs 10, 13). It is interesting that the current fluctuations remained high even at saturating cGMP concentrations (Fig. 3*a*); this perhaps suggests that the binding of cGMP molecules to the channel is an event separate from (though causally related to) the opening and closing of the channel. In other words, a channel molecule can

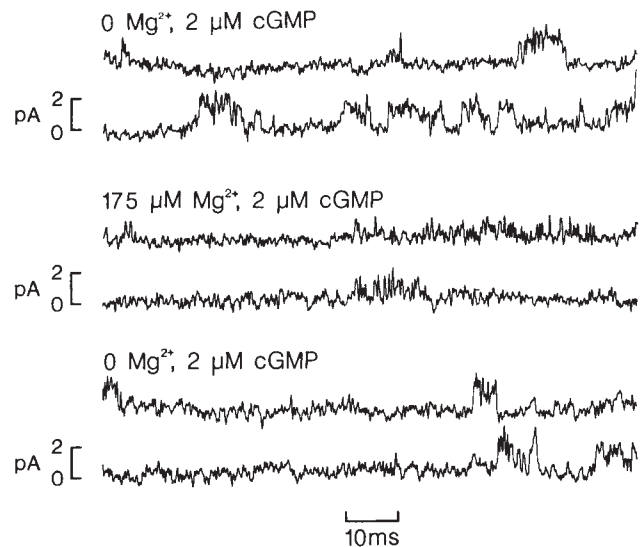


Fig. 4 Flickering block of the open-channel current by Mg^{2+} . Mg^{2+} was added only to the cGMP-containing solution in the bath, while the pipette solution still contained no divalent cations. Holding potential $+60$ mV, salamander rod patch (see Fig. 1 legend). Top two and bottom two traces were obtained with 0 Mg^{2+} ; middle two traces with 175 μM MgCl_2 (no Na-EGTA or Na-EDTA; see Fig. 1 legend). To show the flickering blockade more clearly, the records have been low-pass-filtered at 5 kHz instead of the usual 2 kHz.

open and close repeatedly while still binding cGMP; this same mechanism may also explain the burst phenomenon¹⁵. Figure 3*b* shows the relation between mean current and cGMP concentration based on the records of Fig. 3*a*. The current increased with the third power of cGMP concentration (that is, the Hill coefficient, n , was 3), with a half-saturating cGMP concentration ($K_{1/2}$) of ~ 15 μM . Collected results gave an n of 2.9 ± 0.4 (mean $\pm$ s.d., 15 experiments) and $K_{1/2}$ of 14 – 45 μM at -30 mV (6 experiments). The n value here is higher than previous measurements made in the presence of divalent cations^{1,3,6,7}, but is probably more reliable because the current could be resolved down to lower cGMP concentrations. Still, an influence of divalent cations on the Hill coefficient cannot be ruled out. The $K_{1/2}$ values, on the other hand, are broadly consistent with previous measurements^{1,3,6,7}. There was also an effect of membrane potential on current activation, in that at depolarizing voltages it took a lower concentration of cGMP to elicit a comparable frequency of events (Fig. 3*c*). This voltage effect, however, was not very large, with $K_{1/2}$ decreasing by only $\sim 40\%$ between -40 mV and $+40$ mV (2 experiments), or an estimated e-fold reduction over ~ 160 mV. The Hill coefficient, on the other hand, did not show any obvious voltage dependence. These observations are similar to previous results in cone patches⁸.

Figure 4 shows an example of the effect of divalent cations on the current. With 175 μM Mg^{2+} in the bath, the prominent step changes in the current induced by cGMP were no longer observed. Instead, there were clusters of relatively high-frequency flickers, which most probably reflected blockage of the open channel¹⁶. In other experiments we found that Ca^{2+} had a similar effect, though a higher concentration was required (at least at depolarizing voltages). A flickering block by divalent cations has been observed previously for the glutamate-activated channel in mouse brain¹⁷, as well as for the Ca-channel¹⁸ in conditions where the current is carried by monovalent cations^{19–21}. Indeed, the cGMP-activated (or light-sensitive) conductance has certain properties in common with Ca-channels, in that besides the blockade by divalent cations of current carried by monovalent cations, it is also highly permeable to divalent cations^{22–28}. Both the blockade and the ion selectivity

can perhaps be explained, as has been proposed for the Ca-channel^{21,29}, by the presence of binding sites in the channel with different affinities for monovalent and divalent cations.

Thus our results demonstrate that the light-sensitive conductance is a pore rather than a carrier; this confirms a recent inference from noise measurements on intact cells¹⁰. As the Hill coefficient for current activation is near 3, there must be at least three cGMP binding sites per channel molecule. At present, however, we do not know the details of the kinetic scheme governing the channel activation by cGMP.

We thank Dr Kei Nakatani for invaluable help, Drs Philip Palade and Robert Wong for encouragement and for loaning equipment, and Drs Denis Baylor, Malcolm Brodwick and Philip Palade for discussions. Dr Chris Lingle provided us with an implementation (for the IBM PC) of the single-channel analysis program developed by Dr Frederick Sachs and co-workers. L.W.H. was supported in part by a Fight For Sight Inc. (New York) Research Fellowship; the work was supported by the US National Eye Institute and the Retina Research Foundation, Houston, Texas.

Received 13 December 1985; accepted 4 March 1986.

1. Fesenko, E. E., Kolesnikov, S. S. & Lyubarsky, A. L. *Nature* **313**, 310–313 (1985).
2. Nakatani, K. & Yau, K.-W. *Biophys. J.* **47**, 356a (1985).

3. Yau, K.-W., Haynes, L. W. & Nakatani, K. in *Membrane Control of Cellular Activity* (ed. Lüttgau, H. Ch.) (Gustav Fischer, Stuttgart, in the press).
4. Matthews, G. *Soc. Neurosci. Abstr.* **11**, 1130 (1985).
5. Stern, J. H., Kaupp, U. B. & MacLeish, P. R. *Soc. Neurosci. Abstr.* **11**, 1130 (1985).
6. Zimmerman, A. L., Yamanaka, G., Eckstein, F., Baylor, D. A. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8813–8817 (1985).
7. Yau, K.-W. & Nakatani, K. *Nature* **317**, 252–255 (1985).
8. Haynes, L. W. & Yau, K.-W. *Nature* **317**, 62–65 (1985).
9. Yau, K.-W. & Haynes, L. W. *Biophys. J.* **49**, 33a (1986).
10. Bodoa, R. D. & Detwiler, P. B. *J. Physiol., Lond.* **367**, 183–216 (1985).
11. Lamb, T. D., Matthews, H. R. & Torre, V. J. *J. Physiol., Lond.* **369**, 34P (1985).
12. Gray, P. & Attwell, D. *Proc. R. Soc. B223*, 379–388 (1985).
13. Zimmerman, A. L. & Baylor, D. A. *Biophys. J.* **47**, 357a (1985).
14. Roux, B. & Sauve, R. *Biophys. J.* **48**, 149–158 (1985).
15. Colquhoun, D. & Sakmann, B. *Nature* **294**, 464–466 (1981).
16. Neher, E. & Steinbach, J. H. *J. Physiol., Lond.* **277**, 153–176 (1978).
17. Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462–465 (1984).
18. Lansman, J. B., Hess, P. & Tsien, R. W. *Biophys. J.* **47**, 67a (1985).
19. Kostyuk, P. G., Mironov, S. L. & Shuba, Y. M. *J. Membrane Biol.* **76**, 83–93 (1983).
20. Almers, W., McCleskey, E. W. & Palade, P. T. *J. Physiol., Lond.* **353**, 565–583 (1984).
21. Hess, P. & Tsien, R. W. *Nature* **309**, 453–456 (1984).
22. Capovilla, M., Caretta, A., Cervetto, L. & Torre, V. J. *J. Physiol., Lond.* **343**, 294–310 (1983).
23. Yau, K.-W., McNaughton, P. A. & Hodgkin, A. L. *Nature* **292**, 502–505 (1981).
24. Hodgkin, A. L., McNaughton, P. A., Nunn, B. J. & Yau, K.-W. *J. Physiol., Lond.* **350**, 649–680 (1984).
25. Yau, K.-W. & Nakatani, K. *Nature* **309**, 352–354 (1984).
26. Yau, K.-W. & Nakatani, K. *Nature* **311**, 661–663 (1984).
27. Hodgkin, A. L., McNaughton, P. A. & Nunn, B. J. *J. Physiol., Lond.* **358**, 447–468 (1985).
28. Yau, K.-W. & Nakatani, K. *Nature* **313**, 579–582 (1985).
29. Almers, W. & McCleskey, E. W. *J. Physiol., Lond.* **353**, 585–608 (1984).
30. Sakmann, B. & Neher, E. *Single-Channel Recording* (Plenum, New York, 1983).
31. Sachs, F., Neil, J. & Barkakati, N. *Pflügers Arch. ges. Physiol.* **395**, 331–340 (1982).

Cyclic GMP-sensitive conductance of retinal rods consists of aqueous pores

A. L. Zimmerman & D. A. Baylor

Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305, USA

The surface membrane of retinal rod and cone outer segments contains a cation-selective conductance which is activated by 3',5'-cyclic guanosine monophosphate (cGMP)^{1–4}. Reduction of this conductance by a light-induced decrease in the cytoplasmic concentration of cGMP appears to generate the electrical response to light (reviewed in refs 5, 6), but little is known about the molecular nature of the conductance. The estimated unitary conductance is so small^{1,7–9} that ion transport might occur via either a carrier or a pore mechanism. Here we report recordings of cGMP-activated single-channel currents from excised rod outer segment patches bathed in solutions low in divalent cations. Two elementary conductances, of ~24 and 8 pS, were observed. These conductances are too large to be accounted for by carrier transport, indicating that the cGMP-activated conductance consists of aqueous pores. The dependence of the channel activation on the concentration of cGMP suggests that opening of the pore is triggered by cooperative binding of at least three cGMP molecules.

The cGMP-activated conductance of excised patches from salamander rod outer segments was greatly increased when both sides of the patch were bathed in solutions containing a very low concentration of divalent cations. For five patches in low concentrations of divalent cations (total less than 1 μ M) the maximal agonist-induced conductance at +50 mV was 15.9 ± 2.1 nS (mean $\pm$ s.e.m.), while for 13 patches prepared similarly but studied in solutions with normal levels of divalent cations (external Mg^{2+} concentration ($[Mg]_o$) 6 mM; internal Mg^{2+} concentration ($[Mg]_i$) 3 mM; 1.5 mM $[Ca]_o$) it was 194 ± 27 pS. These results indicate that decreasing the concentration of divalent cations increased the cGMP-sensitive conductance by roughly two orders of magnitude.

Figure 1 shows dose-response relations for the cGMP-sensitive current of patches in 'low divalent' solutions. The cGMP-activated current of each patch is expressed relative to the

maximal value; cGMP concentrations were normalized by the estimated half-saturating concentration, $K_{0.5}$, which ranged from 5.3 to 12 μ M, with an average value of 9.6 μ M. The solid curve was drawn according to the Hill equation: $r/r_{max} = G^3/(1 + G^3)$, where r/r_{max} is the normalized response and G is the concentration of cGMP divided by $K_{0.5}$. This equation would apply if opening of a channel required the binding of at least three molecules of cGMP, and if the binding were strongly cooperative. The dashed curve, which gives a poorer fit, was drawn according to $r/r_{max} = (G/[1 + G])^3$; this relation is expected if opening requires binding to three independent sites of equal affinity. Somewhat smaller power exponents have been reported from experiments with higher concentrations of divalent cations^{1,3,10,11}, where the smaller size of the agonist-induced currents prevented measurement of the foot of the dose-response relation.

Figure 2 shows the current obtained from an excised patch at a holding potential of +50 mV. Application of 0.5 or 1 μ M cGMP to the cytoplasmic side of the patch caused outward current pulses, which were much more frequent at the higher dose. Removal of cGMP abolished the pulses, demonstrating that they did not result from disruption of the membrane or seal.

Selected records from the experiment of Fig. 2 are plotted on an expanded time base in Fig. 3A, which shows unitary outward currents with a mean amplitude close to 1 pA and widely variable durations. For the patch potential of +50 mV the underlying conductance change is 20 pS. In some traces there was a suggestion of smaller events of amplitude ~0.3 pA (conductance = 6 pS). Such smaller events are better resolved in the experiment of Fig. 4. Similar experiments on a total of nine patches gave, for the large and small transitions, mean conductance changes of 23.7 ± 4.1 and 7.7 ± 1.4 pS (mean $\pm$ s.d.), respectively. In one experiment the amplitude of the single-channel currents varied approximately linearly with holding potential between -100 and +50 mV.

Figure 3B presents further evidence for the quantal nature of the current transitions in the experiment of Fig. 2. The control amplitude histogram (Fig. 3B, a) is approximately gaussian, with a width at half height of 0.3 pA. The histogram for 0.5 μ M cGMP (Fig. 3B, b) shows a small additional hump between 0.3 and 1 pA. In 1.0 μ M cGMP (Fig. 3B, c) there are two peaks centred at amplitudes of 0 and 0.8 pA, with a smaller hump at ~1.6 pA. The form of the three histograms is consistent with quantal agonist-induced currents of ~0.8 pA, like those in

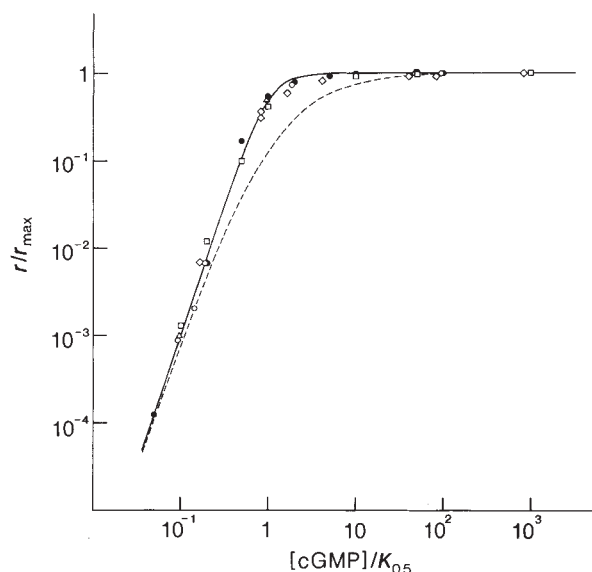


Fig. 1 Dose-response relation for activation of the cGMP-sensitive conductance of excised rod outer segment patches in low divalent solutions. Results from five patches, each plotted using a different symbol. The ordinate is the cGMP-activated current (r) relative to the maximal value ($r_{\max}$), while the abscissa is the cGMP concentration relative to the estimated half-saturating concentration $K_{0.5}$. The curves were calculated from equations described in the text. Saturating currents and values of $K_{0.5}$ for each patch were as follows: $\circ$: 1,021 pA, 5.3 μ M; $\triangle$: 116 pA, 10.5 μ M; $\diamond$: 834 pA, 12 μ M; $\square$: 774 pA, 10.0 μ M; $\bullet$: 1,234 pA, 10.0 μ M.

Methods. Patches were excised from the distal tip of isolated salamander rod outer segments, which were prepared in the light¹ and cleaned by exposure for 5 min to ~ 0.5 mg ml⁻¹ hyaluronidase (Cooper, HSEP, 5474). Patches were studied in symmetrical solutions consisting of either nominally 0 divalent Ringer's, containing (in mM): NaCl 111, KCl 2.5, EDTA 0.02, pH 7.6 HEPES buffer 3.0, dextrose 10, sucrose 22.5 (solution A); or NaCl 128.5, pH 7.6 HEPES buffer 2.0, MgCl₂ 0.02, EDTA 0.20 (solution B), calculated to give a free Mg concentration of 10^{-7} M and a free Ca concentration $< 10^{-7}$ M. cGMP was applied to the cytoplasmic side of the patches by perfusion through the recording chamber. The patch potential was held at +50 mV (cytoplasmic side positive), and the cGMP-activated current determined as the difference between the currents in the presence and absence of cGMP.

Fig. 3A. Analysis of histogram *c* in Fig. 3B showed that the valley between the main peaks was higher than expected from the summation of two gaussians, perhaps because of the presence of the smaller events. Furthermore, the histogram was not well fitted by a Poisson distribution constructed on the assumption that there were only large events and that they summed linearly; based on this model more samples are expected near an amplitude of $2 \times 0.8 = 1.6$ pA, given the small magnitude of the peak near 0 pA.

Analysis of the fluctuations in the macroscopic cGMP-activated current provided evidence that single-channel currents like those described above sum to produce the current at higher concentrations of cGMP. In four experiments the apparent single-channel current amplitude $\bar{a}$ was estimated from the slope of the relation between the mean response r and the scaled cGMP-induced variance, $\sigma^2 r_{\max}/(r_{\max} - r)$, which was determined over a bandwidth of only 100 Hz in order to minimize high-frequency noise arising from fast conductance fluctuations in activated channels. For values of r up to half-saturating, $\bar{a}$ was 0.46 ± 0.07 pA (mean $\pm$ s.d.) at +50 mV. This value of $\bar{a}$ falls between the mean amplitudes of the large and small transitions that were measured directly at low cGMP concentrations.

Single-channel currents of 0.3 or 1 pA are 100 times larger than those mediated by a variety of membrane carriers¹², sug-

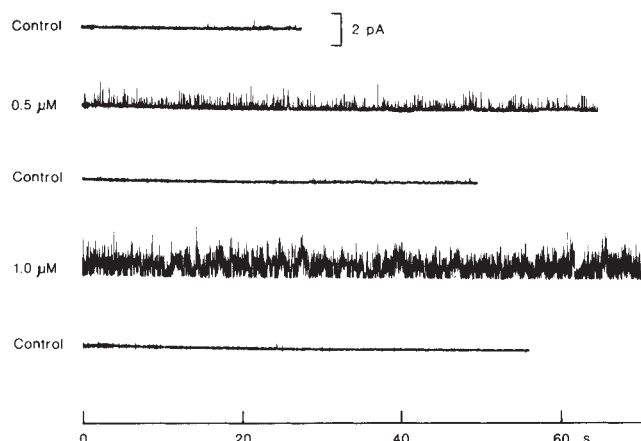


Fig. 2 Currents from an excised patch during application of low concentrations of cGMP to the cytoplasmic side. The patch was in the low divalent solution (A) of Fig. 1 legend. The recording is continuous except for interruptions at the ends of the traces, where the chamber was perfused for solution changes. Patch potential +50 mV, outward change in membrane current plotted upward. Bandwidth 0–100 Hz, temperature 21.8 °C, leakage resistance 10 G Ω .

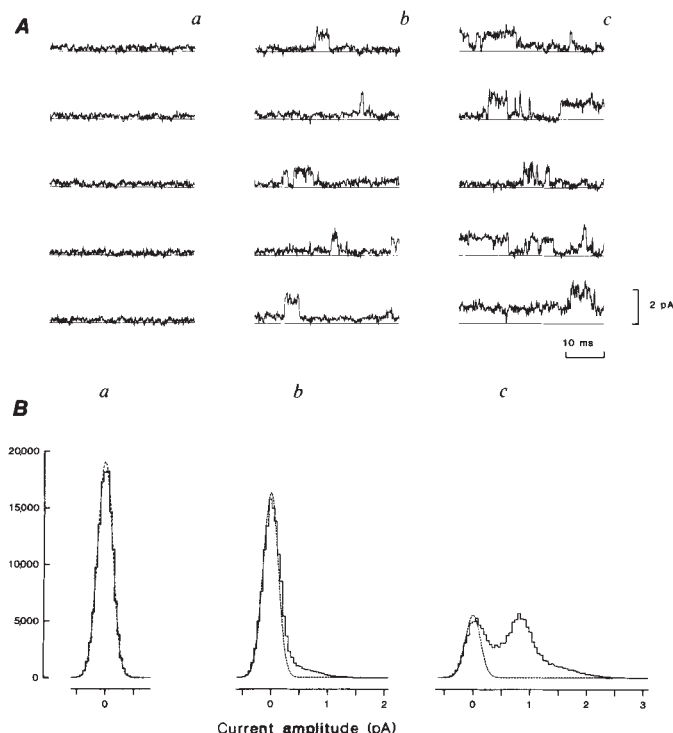


Fig. 3 Characteristics of the cGMP-induced patch currents from the experiment of Fig. 2. **A**, Sample patch currents from control, 0.5 μ M cGMP and 1.0 μ M cGMP runs (*a–c*, respectively). Lines beneath the traces were drawn at the same absolute current level. Recordings were high-pass-filtered with a single pole filter of time constant 10 s to remove a steady d.c. drift of 0.003 pA s⁻¹. Low-pass filtering was at 2 kHz (–3 dB point, 8-pole Bessel filter; Frequency Devices). Each sweep consists of 512 points acquired at a sampling rate of 10 kHz. **B**, Amplitude histograms from the experiment of Fig. 2 (*a–c* same as for **A**). Each histogram was prepared from 250 sweeps filtered and sampled as in **A**. The control histogram (*a*) was constructed from 50 sweeps from the initial control run and 100 sweeps from each of the later control runs of Fig. 2. Dashed curves are gaussian distributions with standard deviation $\sigma = 0.134$ pA, the measured value of r.m.s. noise in the control sweeps. The scaling of the gaussians corresponds to baseline probabilities, P_0 , of 0.855 (0.5 μ M) and 0.289 (1 μ M). The average number of open channels, calculated from $-\ln P_0$, was 0.157 (0.5 μ M) and 1.24 (1.0 μ M).

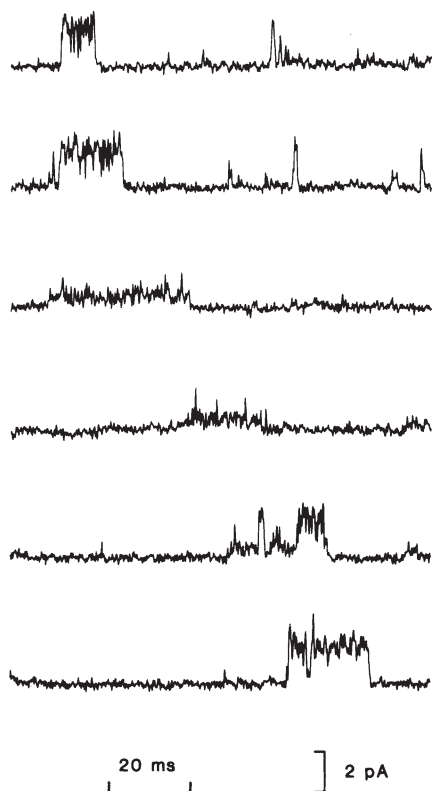


Fig. 4 Evidence for two activated states having different mean conductances. The currents are from an excised patch exposed to $0.5 \mu\text{M}$ cGMP; holding potential $+75 \text{ mV}$, bandwidth 0.2 kHz , sampling rate 10 kHz . The patch was in solution B of Fig. 1 legend. Leakage resistance $10 \text{ G}\Omega$.

gesting a pore mechanism of conduction. The effect of divalent cations in reducing the conductance might be explained by assuming that ions move through the channel in single file and encounter one or more binding sites with a high affinity for divalent cations. Divalent cations would pass through the channel^{13–15}, but with a long transit time, giving a very small time-averaged single-channel current under normal conditions⁸.

The total number of channels in a patch can be estimated from the ratio of the current at saturating cGMP concentration to the single-channel current. In the experiments of Fig. 1, the saturating currents averaged 800 pA (range $116\text{--}1,234 \text{ pA}$). Taking the single-channel current as 1 pA and the patch area as roughly $2 \mu\text{m}^2$, the apparent channel density is $\sim 400 \mu\text{m}^{-2}$. A channel density of $200 \mu\text{m}^{-2}$ was estimated from an analysis of the light-sensitive current noise of whole rods⁹. These rough estimates seem to be in satisfactory agreement, supporting the identity of the light-sensitive and cGMP-sensitive channels.

The large and small conductance states are illustrated in the records of Fig. 4, obtained from a patch held at $+75 \text{ mV}$. The upper two traces show outward currents of $\sim 1.9 \text{ pA}$ mean amplitude (conductance 25 pS), while the middle two traces show smaller events with a mean amplitude of $\sim 0.7 \text{ pA}$ (conductance 9 pS). Both types of currents were noisy, suggesting that there were conductance fluctuations so fast that the method did not resolve them. The two lower traces in Fig. 4 illustrate events suggestive of transitions between the small- and large-conductance states: the highest current level is comparable to the large amplitude alone, rather than to the sum of the large and small amplitudes. It remains to be determined whether the two states have the same peak instantaneous conductance.

We conclude that the cGMP-activated channel of retinal rods is an aqueous pore which has a very small effective conductance in the presence of physiological concentrations of divalent

cations. The small conductance is appropriate to give low quantizing noise in the small, stereotyped signals that encode photon absorption¹⁶. The cubic dependence of channel opening on cGMP concentration makes the channel an efficient molecular switch which is very sensitive to the local cGMP concentration.

We thank Dr K.-W. Yau for suggestions on divalent cation effects, Drs L. Stryer and R. Aldrich for discussions, and Mr R. Schneeveis for technical help. This work was supported by grant EY01543 from the National Eye Institute, USPHS.

Received 4 December 1985; accepted 6 March 1986.

1. Fesenko, E. E., Kolesnikov, S. S. & Lyubarsky, A. L. *Nature* **313**, 310–313 (1985).
2. Nakatani, K. & Yau, K.-W. *Biophys. J.* **47**, 365a (1985).
3. Yau, K.-W. & Nakatani, K. *Nature* **317**, 252–255 (1985).
4. Haynes, L. & Yau, K.-W. *Nature* **317**, 61–64 (1985).
5. Altman, J. *Nature* **313**, 264–265 (1985).
6. Stryer, L. *A. Rev. Neurosci.* **9**, 87–119 (1986).
7. Gray, P. & Attwell, D. *Proc. R. Soc. B* **223**, 379–388 (1985).
8. Bodia, R. D. & Detwiler, P. B. *J. Physiol., Lond.* **367**, 183–216 (1985).
9. Zimmerman, A. L. & Baylor, D. A. *Biophys. J.* **47**, 357a (1985).
10. Koch, K.-W. & Kaupp, U. B. *J. biol. Chem.* **260**, 6788–6800 (1985).
11. Zimmerman, A. L., Yamanaka, G., Eckstein, F., Baylor, D. A. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8813–8817 (1985).
12. Hille, B. *Ionic Channels of Excitable Membranes*, 202 (Sinauer Associates, Sunderland, Massachusetts, 1984).
13. Yau, K.-W., McNaughton, P. A. & Hodgkin, A. L. *Nature* **292**, 502–505 (1981).
14. Capovilla, M., Caretta, A., Cervetto, L. & Torre, V. *J. Physiol., Lond.* **343**, 295–310 (1983).
15. Hodgkin, A. L., McNaughton, P. A. & Nunn, B. J. *J. Physiol., Lond.* **358**, 447–468 (1985).
16. Baylor, D. A., Lamb, T. D. & Yau, K.-W. *J. Physiol., Lond.* **288**, 613–634 (1979).

Neural control of the forms of acetylcholinesterase in slow mammalian muscles

Josephine Lai*, Jolanta Jedrzejczyk†, John A. Pizzey‡, David Green & Eric A. Barnard§

Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, and MRC Molecular Neurobiology Unit, MRC Centre, Hills Road, Cambridge CB2 2QH, UK

The 'heavy', collagen-tailed form of acetylcholinesterase (AChE), having a $s_{20,w}^0$ of 16S in mammals, occurs at vertebrate muscle endplates and has been widely regarded as a marker of neuronal influence on muscle *in vivo*^{1–5}. However, an interesting exception has been described by Bacou *et al.*, in a previous report in *Nature*⁶. They found, in a slow-twitch muscle of the rabbit, that after denervation the 16S form of AChE increases markedly, rather than disappearing. Such a phenomenon would modify current concepts of neuromuscular regulation. We report here, however, that this exception is apparent rather than real in terms of endplate AChE regulation.

The rabbit semimembranosus muscle was described by Bacou *et al.*⁶ as containing an 'external region I' composed of fast-twitch glycolytic muscle (type IIB), surrounding an 'internal conoidal region II' containing only slow-twitch (type I) fibres. In fact, these prominent rabbit leg muscles have previously been named, respectively, as the semimembranosus accessorius (SA) and the semimembranosus proprius (SP) muscles^{7,8}; the SA has been shown in rabbits to contain 86% type IIB, 12% type IIA and 2% type I fibres⁹, while the SP contains 100% type I fibres. These values hold in our rabbits.

The AChE concentration in the normal rabbit SP (Fig. 1a) differs in our analyses from that reported⁶. The 16S form is major, not minor, and the light forms are mainly the 6S (dimeric),

* Present address: Laboratory of Neurophysiology, NINCOS, NIH, Bethesda, Maryland 20205, USA.

† Permanent address: Department of Histology, Academy of Medicine, Warsaw 02-004, Poland.

‡ Present address: Biophysics Department, KQC Kensington Campus, London W8 7AH, UK.

§ To whom correspondence should be addressed at Cambridge.

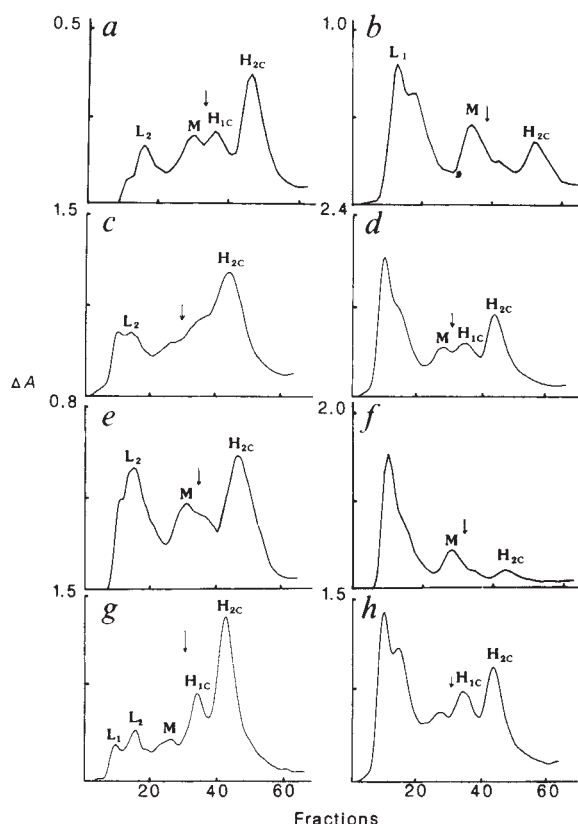


Fig. 1 Sucrose density gradient profiles of AChE in control and chronically denervated (4 weeks) muscles. Right-hand panels show denervated muscles, left-hand panels show control muscles. *a, b*, Rabbit SP; *c, d*, guinea pig SP; *e, f*, rabbit soleus; *g, h*, guinea pig soleus. The amounts of control tissue taken were 1.3–1.7 times those of the denervated tissues. M denotes the presumed tetrameric, globular form and H_{1c} the smaller collagen-tailed form, ~13S, which has fewer catalytic units than the maximum in the H_{2c} (~16S) form¹. L_1 and L_2 denote, respectively, the ~4S and ~6S forms. AChE activity is in units of absorbance change (ΔA)¹ (note the different scales). The total centrifugal force used varied: arrows show the catalase (11.4S) internal marker.

not the 4S (monomeric), AChE. Equivalent differences in the SA were also found (Table 1). We attribute these differences to our use of more antiproteases (see Fig. 1 legend) and to particularly rapid processing. When these precautions were relaxed, profiles similar to those reported⁶ were obtained. High 4S AChE in mature muscles is generally a marker of proteolytic breakdown^{1,4,10}. Hence, the starting concentration of SP 16S AChE was much higher than that (only 9.5% of the total AChE) found previously⁶. After complete denervation (see below) for 2–4 weeks, the contribution of the 16S form to the total AChE declined greatly in both muscles (Fig. 1*b*, Table 1). Similar behaviour was observed in other denervated slow-twitch muscles, that is, the SP of the guinea pig (which we have identified as being similar anatomically and in fibre types to rabbit SP), as well as rabbit and guinea pig soleus muscles (Fig. 1*c–h*).

It has been frequently reported that mammalian muscles lose nearly all of their 16S AChE when the maximum effect of denervation is attained^{3,5,11}. We found, however, that although there was a decline in the 16S form, some persisted throughout, with higher concentrations remaining in certain slow-twitch muscles (Fig. 1, Table 1). Further, on chronic denervation, the AChE concentration (and the AChE content per entire muscle) may either decrease (the most usual case) or increase, depending

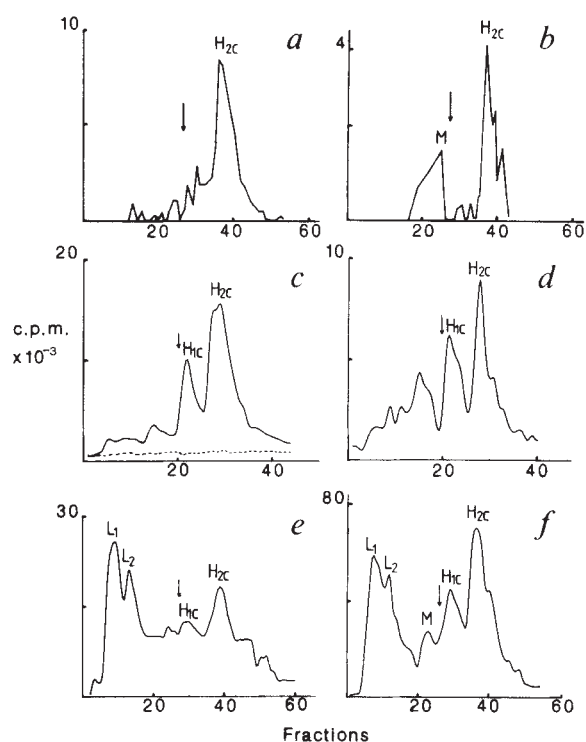


Fig. 2 Sucrose density gradient analysis of AChE (fractions assayed on 3H -acetylcholine)¹⁵ in the endplate (left-hand panels) and the non-endplate (right-hand panels) regions of slow-twitch muscles. *a, b*, Rabbit SP; *c, d*, guinea pig soleus; *e, f*, denervated guinea pig soleus. Endplates and non-endplate segments were microdissected¹⁵ from the same fibres; 20–50 endplates (with the minimum of underlying fibre) or the same number of non-endplate segments (the latter about five times the length of an endplate and cut transversely across the whole fibre) were extracted for 60–90 min at room temperature in 0.25 ml of the extraction buffer described in Table 1 legend and all was applied. A blank was run in parallel with an identical extract, and was assayed in the presence of 10^{-4} M isooctamethyl pyrophosphoramidate and 5×10^{-5} M BW284C51 (see ref. 15) to inhibit all cholinesterases; this is illustrated by the broken line in *c*, while in *a, b, e, f* the blank has already been subtracted throughout. The total force and the fraction size used varied, the arrow shows the catalase marker, and only the horizontal pairs can be compared in the amounts. Labelling of peaks is as in Fig. 1. Note that the smaller collagen-tailed form, H_{1c} , is also prominent in the guinea pig muscle both in and outside the endplate. Similar profiles were found when repeated on other guinea pigs and rabbits: see Table 1 for the means of the contributions of the L and H_{2c} peaks in all the experiments.

on the species; rabbit¹² and chicken¹⁰ were previously known to behave exceptionally in increasing their muscle AChE. We find that the guinea pig is also in that group. In those three species a long-lasting increase of severalfold occurred in the AChE concentration in both fast and slow muscles after denervation, and despite the muscle weight loss (to the order of one-half, at the maximum here) the total content of AChE per muscle did not decline but increased further. Nevertheless, the proportion of 16S AChE in the slow muscles was much reduced, while the 16S AChE per entire muscle remained about constant or again decreased (Table 1). The overall increase in AChE was contributed by the lighter forms. In the fast fibres the 16S AChE per entire muscle always declined greatly on denervation.

Another technical reason for the difference in findings may be the surgical procedure involved. We fully denervated both SP and SA, by cutting all relevant branches of the sciatic and obturator nerves (a branch of the latter giving an additional

Table 1 Changes in AChE activities in slow- and fast-twitch muscles on denervation

Species	Muscle	Control [AChE] (U g ⁻¹)	n	Ratio of denervated to control			L (%)	16S (%)
				(basis)	all forms	L		
Rabbit	SP (100% slow)	0.6 ± 0.07	7	(total)	1.7 ± 0.3	2.7	55(16)	19(54)
				(conc.)	4.4 ± 0.7	7.0		
	Soleus (96% slow)	0.6 ± 0.07	6	(total)	1.3 ± 0.3	3.1	72(33)	10(42)
Guinea pig				(conc.)	2.5 ± 0.4	5.8		
	SA (98% fast)	0.5 ± 0.03	7	(total)	1.6 ± 0.1	4.4	73(26)	10(44)
				(conc.)	3.7 ± 0.8	10.3		
	SP (100% slow)	0.7 ± 0.04	5	(total)	1.7 ± 0.2	3.0	48(27)	25(43)
				(conc.)	3.5 ± 0.3	6.2		
	Soleus (100% slow)	0.5 ± 0.05	6	(total)	1.5 ± 0.3	2.0	44(17)	25(71)
				(conc.)	3.0 ± 0.4	4.0		
	SA (99% fast)	0.6 ± 0.01	5	(total)	1.2 ± 0.1	2.8	60(32)	13(44)
				(conc.)	2.7 ± 0.4	6.2		
	EDL (94% fast)	0.8 ± 0.02	3	(conc.)	1.3 ± 0.2	2.2	62(30)	14(51)

Values given are the ratios of the activity of specified AChE forms in the denervated muscles relative to that of the corresponding activity in the contralateral control muscles: the ratio of the activities is expressed either on the basis of per gram (wet weight) of muscle ('conc.') or (for the same specimens) on the basis of the entire muscle content ('total'). The mean ± s.e.m. for *n* denervated animals sampled is given for each muscle type. (Where s.e. is not shown, it was of the same relative order as given for that muscle or ratio.) The last two columns show the % of the muscle AChE of the denervated muscle (or, in parentheses, of the control muscle) contributed by all the L (~4S and ~6S) forms, or by the 16S form. The difference of their sum from 100% is all accounted for by the forms sedimenting in the region of the M (10S) and H_{1c} (13S) peaks (see Fig. 1), whose poor resolution frequently prevents accurate computation. The fibre typing noted shows the % of type I fibres (slow) or type II fibres (fast) in these muscles. The soleus or extensor digitorum longus (EDL) muscles of New Zealand White rabbits (~3 kg) or Dunkin-Hartley guinea pigs (2 months old) were denervated by sectioning the lower branch of the sciatic nerve. The SP of the rabbit was fully denervated by sectioning the branches of the sciatic nerve, the cutaneous caudal femoral nerve and the obturator nerve to that muscle⁸. Similarly, the guinea pig SP was denervated by cutting the branches of the sciatic nerve, the cutaneous caudal femoral nerve, the femoral nerve and the saphenous nerve to that muscle. The rest of the limb remained mobile in each case. The control (contralateral) and denervated muscles were analysed for their AChE content 4 weeks later. Changes in the same direction, but to a lesser extent, were seen in other cases analysed after 2 or 3 weeks of denervation. For AChE analysis, muscle crude extracts were processed rapidly at 4 °C: the entire muscle was carefully excised, weighed, minced rapidly, mixed and aliquots were homogenized in 10 vol. of 5 mM EGTA, bacitracin (1 mg ml⁻¹), benzamidinium-HCl (2 mM), *N*-ethylmaleimide (5 mM), chicken ovomucoid inhibitor (0.2 mg ml⁻¹), soybean trypsin inhibitor (0.1 mg ml⁻¹), leupeptin (0.04 mg ml⁻¹) and pepstatin (0.02 mg ml⁻¹). The homogenates were centrifuged at 43,500g for 15 min, and aliquots of the extracts thus obtained were used immediately for sucrose density gradient analysis (Fig. 1) or AChE assays¹. The proportions of the molecular forms were obtained from computer-assisted analysis of the peaks in the profile.

motor supply). Bacou *et al.*⁶ cut only the sciatic nerve. When we did this, only partial denervation occurred; a twitch of the semimembranosus could still be elicited by stimulating the obturator nerve. Whether denervation was partial or total, the morphological changes were similar and as illustrated by Bacou *et al.*⁶: the SA was more affected than the SP muscle, with marked fibre atrophy in both, and similarly in other denervated slow muscles such as rabbit and guinea pig soleus. Despite the similar histology, partial denervation may produce increased activity or terminal sprouting¹³.

Further, we localized the various AChE forms, using microdissection^{14,15} of individual endplates (Fig. 2). In the innervated slow-twitch fibres, the endplate contained essentially only collagen-tailed (16S and a little 13S) AChE (Fig. 2*a,c*). However, the extrasynaptic regions of innervated slow-twitch muscles studied also contained most of their AChE in these heavy forms (Fig. 2*b,d*). Although at a much lower concentration than at the endplate, when summed over the very great total surface this forms a large contribution to the whole muscle AChE. This extrasynaptic collagen-tailed AChE (also seen in non-endplate segments of rat diaphragm¹¹, where 40% of fibres are type I, and in hamster slow-twitch fibres) is much lower or absent in fast-twitch fibres^{4,14}. The microdissected extrasynaptic regions of the denervated slow-twitch fibres of the SP and soleus showed a large increase in the light forms but also clear retention of the heavy forms (Fig. 2*f*); their endplates showed much loss of the 16S and the appearance of 4S and 6S AChE (Fig. 2*e*).

We conclude that denervation always produces a decline in the proportion of the muscle collagen-tailed AChE and in its

amount at the endplates. However, in slow-twitch fibres there is, interestingly, another population of extrasynaptic 16S AChE which does not require neural contact. In a few species where AChE production increases on denervation, this extrasynaptic 16S AChE increases with the rest; this gain is nevertheless balanced or exceeded by the loss from the endplates, and the percentage contribution of this form to the total AChE of the muscle invariably declines markedly.

J.L. held a Croucher Foundation Graduate Studentship and J.J. held a Wellcome Trust Fellowship. We thank the Muscular Dystrophy Group of Great Britain for additional support.

Received 27 January; accepted 10 March 1986.

1. Lyles, J. M., Silman, I., Di Giambardino, L., Couraud, J. Y. & Barnard, E. A. *J. Neurochem.* **38**, 1007-1021 (1982).
2. Massoulie, J. M. & Bon, S. A. *Rev. Neurosci.* **5**, 57-106 (1982).
3. Hall, Z. W. *J. Neurobiol.* **4**, 343-361 (1973).
4. Randall, W. R., Lai, J. & Barnard, E. A. in *Molecular Basis of Nerve Activity* (eds Changeux, J.-P., Hucho, F., Maelicke, A. & Neumann, E.) 595-618 (de Gruyter, Berlin, 1985).
5. Vigny, M., Koenig, J. & Rieger, F. *J. Neurochem.* **27**, 1347-1353 (1976).
6. Bacou, F., Vigneron, P. & Massoulie, J. *Nature* **296**, 661-664 (1982).
7. Kerr, N. S. *Anat. Rec.* **121**, 481 (1955).
8. Narone, R., Pavaux, C., Blin, P. C. & Cuq, P. *Atlas D'Anatomie du Lapin* (Masson, Paris, 1973).
9. Peter, J. B., Barnard, P. J., Edgerton, V. R., Gillespie, C. A. & Stempel, K. E. *Biochemistry* **11**, 2627-2633 (1972).
10. Silman, I., Di Giambardino, L., Lyles, J. M., Couraud, J. Y. & Barnard, E. A. *Nature* **280**, 160-162 (1980).
11. Collins, P. L. & Younkin, S. G. *J. Biol. Chem.* **257**, 13638-13644 (1982).
12. Tennyson, V. M., Kremzner, L. T. & Brzin, M. *J. Neuropath. exp. Neurol.* **36**, 245-275 (1977).
13. Betz, W. J., Caldwell, J. H. & Ribchester, R. R. *J. Physiol., Lond.* **303**, 265-280 (1980).
14. Jedrzejczyk, J., Silman, I., Lyles, J. M. & Barnard, E. A. *Biosci. Rep.* **1**, 45-51 (1981).
15. Jedrzejczyk, J., Silman, I., Lai, J. & Barnard, E. A. *Neurosci. Lett.* **46**, 283-289 (1984).

Cloning of the gene and cDNA for mammalian β -adrenergic receptor and homology with rhodopsin

Richard A. F. Dixon*, Brian K. Kobilka†, David J. Strader‡, Jeffrey L. Benovic†, Henrik G. Dohlman†, Thomas Frielle†, Mark A. Bolanowski†, Carl D. Bennett§, Elaine Rands*, Ronald E. Diehl*, Richard A. Mumford‡, Eve E. Slater‡, Irving S. Sigal*, Marc G. Caron†, Robert J. Lefkowitz† & Catherine D. Strader‡

Departments of *Virus and Cell Biology Research and §Medicinal Chemistry, Merck Sharp and Dohme Research Laboratories, West Point, Pennsylvania 19486, USA

†Howard Hughes Medical Institute, Department of Medicine, Biochemistry and Physiology, Duke University Medical Center, Durham, North Carolina 27710, USA

‡Department of Biochemistry and Molecular Biology, Merck Sharp and Dohme Research Laboratories, Rahway, New Jersey 07065, USA

The adenylate cyclase system, which consists of a catalytic moiety and regulatory guanine nucleotide-binding proteins, provides the effector mechanism for the intracellular actions of many hormones and drugs¹. The tissue specificity of the system is determined by the particular receptors that a cell expresses. Of the many receptors known to modulate adenylate cyclase activity, the best characterized and one of the most pharmacologically important is the β -adrenergic receptor (β AR). The pharmacologically distinguishable subtypes of the β -adrenergic receptor, β_1 and β_2 receptors, stimulate adenylate cyclase on binding specific catecholamines¹. Recently, the avian erythrocyte β_1 , the amphibian erythrocyte β_2 and the mammalian lung β_2 receptors have been purified to homogeneity and demonstrated to retain binding activity in detergent-solubilized form¹⁻⁵. Moreover, the β -adrenergic receptor has been reconstituted with the other components of the adenylate cyclase system *in vitro*⁶, thus making this hormone receptor particularly attractive for studies of the mechanism of receptor action. This situation is in contrast to that for the receptors for growth factors and insulin, where the primary biochemical effectors of receptor action are unknown. Here, we report the cloning of the gene and cDNA for the mammalian β_2 AR. Analysis of the amino-acid sequence predicted for the β AR indicates significant amino-acid homology with bovine rhodopsin and suggests that, like rhodopsin⁷, β AR possesses multiple membrane-spanning regions.

Hamster lung β AR was purified to homogeneity by sequential affinity chromatography and molecular-sieve HPLC as described previously^{2,5}. The purified receptor bound ligand with theoretical specific activity and migrated on SDS-polyacrylamide gel electrophoresis as a single broad band at a relative molecular mass (M_r) of 64,000 (64K). Initial attempts to obtain N-terminal sequence data on intact β AR failed, presumably because the N-terminus of this protein was blocked. Therefore, peptide fragments generated by CNBr cleavage of pure β AR were isolated by reverse-phase HPLC.

Figure 1a (solid line) shows a peptide map generated from 1 nmol of pure receptor; the broken line shows the HPLC profile resulting from CNBr treatment of the detergent alone. The β AR-derived peptides produced at least nine specific absorbance peaks which were reproducibly observed in five separate β AR preparations. The most prominent of these peptides (marked with arrows in Fig. 1) were subjected to N-terminal sequence analysis, yielding the amino-acid sequences in Fig. 1b.

To confirm that the determined amino-acid sequences were those of the β AR polypeptide, we raised anti-peptide antibodies against peptide 7. Peptide 7 was expressed in *Escherichia coli* as a C-terminal peptide fused to the N-terminal domain of the

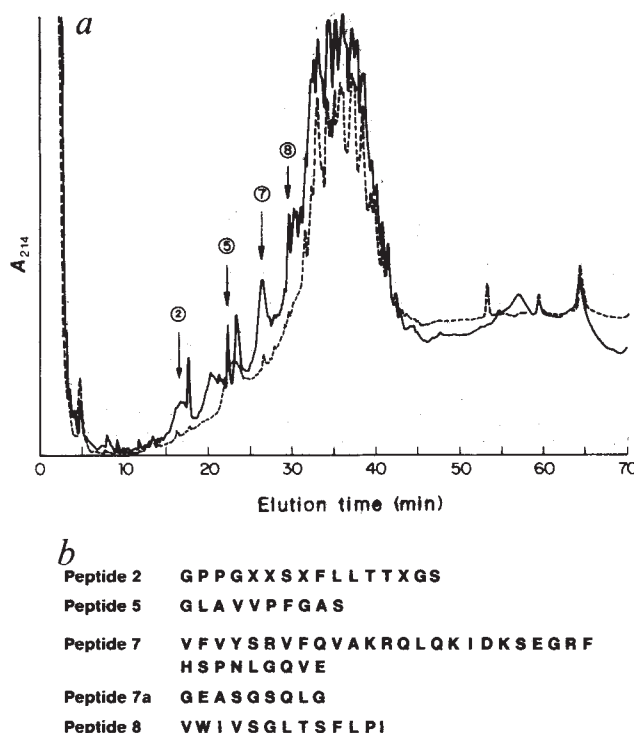


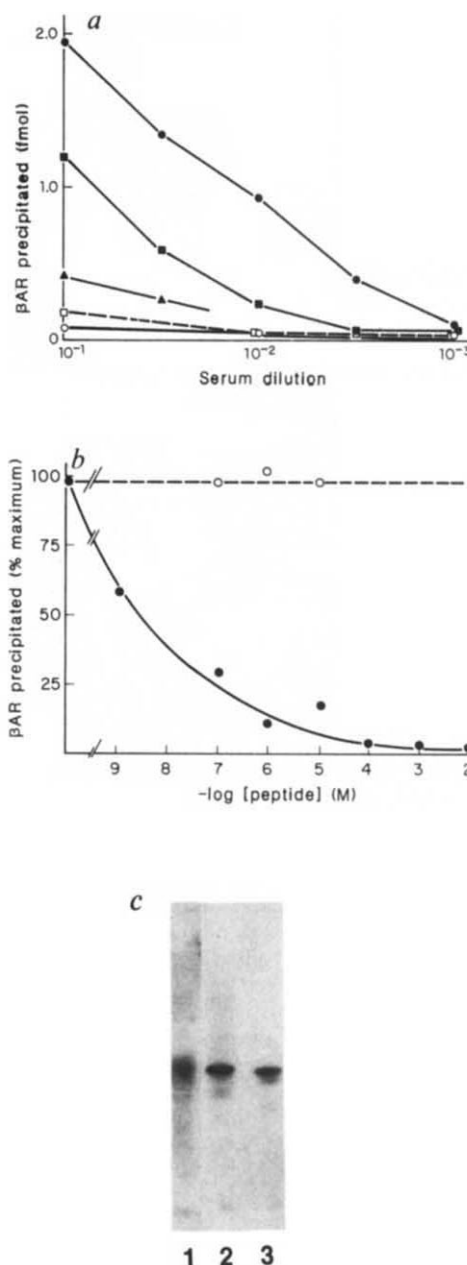
Fig. 1 Amino-acid sequence of peptides derived from CNBr-treated β -adrenergic receptor. **a**, Absorbance profiles represent CNBr treatment of pure β AR (solid line) or digitonin (dashed line). The arrows indicate the peptides that were sequenced. **b**, Amino-acid sequences identified by HPLC following each cycle of the sequenator. Two of the four blank cycles (X) in the amino-acid sequence for peptide 2 are presumed to be due to N-linked glycosylation. Peptides 7 and 7a were located within the same peak. **Methods.** β AR was purified to homogeneity from hamster lung membranes by the method of Benovic *et al.*⁵, using affinity chromatography followed by molecular-sieve HPLC. Binding of ¹²⁵I-CYP to intact cells or to solubilized β AR was determined according to Caron and Lefkowitz²². For peptide preparation, ~1 nmol of pure β AR was treated with CNBr (0.4 mM) in 70% formic acid at 23 °C for 20 h. After lyophilization, the sample was resuspended in 20 mM trifluoroacetic acid (TFA) and the peptides separated by reverse-phase HPLC on a Synchropak C-4 column, eluted with a 10–70% acetonitrile gradient containing 20 mM TFA. The N-terminal sequence analysis was performed by the method of Hewick *et al.*²³, using a gas-phase sequenator (Applied Biosystems). The phenylthiohydantoin (PTH) amino acids produced at each step were separated and quantitated by HPLC²⁴.

yeast RAS^{sc1} protein SC1N (ref. 8). Rabbits injected with the isolated fusion protein produced antibodies which reacted with ¹²⁵I-labelled immunogen as well as pure ¹²⁵I-labelled β AR or pure ¹²⁵I-cyanopindolol-labelled (¹²⁵I-CYP) β AR. Figure 2a shows an immunoprecipitation titration curve of this antibody against ¹²⁵I-CYP-labelled solubilized β AR from hamster lung, hamster heart, A431 epidermoid carcinoma cells and turkey erythrocytes. No immunoprecipitation of counts above background was observed in control experiments when ¹²⁵I-CYP was incubated with the antibody in the absence of receptor.

Antibody to the hamster lung β AR-derived peptide was capable of recognizing the human β AR from the A431 line, albeit with a slightly lower sensitivity (50%) (Fig. 2a). The antibody also cross-reacted slightly with hamster heart β AR, a tissue containing a β_1 subtype of receptor, but did not immunoprecipitate the β_1 AR of turkey erythrocytes. These differences in antibody sensitivity could reflect differences in either primary sequence or the conformation of this region of the protein within the various receptor subtypes and receptors from different species. To confirm that the antibody was recog-

Fig. 2 Immunoreactivity of β -adrenergic receptor. *a*, Immunoprecipitation of 125 I-CYP- β AR from hamster lung (●), hamster heart (▲), A431 cells (■) or turkey erythrocytes (○) by serum from rabbits immunized with β AR peptide 7. □, Immunoprecipitation of hamster lung receptor by preimmune serum. *b*, Immunoprecipitation of 125 I-CYP-labelled hamster lung β AR with anti- β AR peptide 7 antibody following preincubation with a synthetic peptide containing a portion of the sequence for peptide 7 (●), or with an unrelated peptide, atrial natriuretic factor (○). *c*, Protein immunoblotting of the β AR. Protein samples were separated on a 10% polyacrylamide gel²⁵, transferred to nitrocellulose and treated sequentially with antibodies and 125 I-protein A (10^6 c.p.m. per 25 ml) as described elsewhere^{26,27}. Lane 1, pure hamster lung β AR (5 pmol); lane 2, A431 cell lysate from 4×10^5 cells; lane 3, RPMI 1846 lysate from 10^5 cells.

Methods. The immunogen for induction of anti-peptide antibodies was expressed in *E. coli*. Two oligonucleotides encoding the 19 amino acids of peptide 7 (QVAKRQLQKIDKSEGRFHS) were synthesized (see Fig. 1 legend), annealed and ligated²⁸ into the *AccI* and *HindIII* sites of plasmid pSC1N⁸ to give the plasmid p β P1. *E. coli* transformed with p β P1 overexpresses a protein of apparent M_r 23K, while *E. coli* containing pSC1N overexpresses a protein of apparent M_r 21K (data not shown). This observed difference in relative molecular mass of the two proteins is consistent with the encoded fusion protein containing 19 additional amino acids. To prepare antigen, plasmid-containing cells were grown in L-broth containing ampicillin and isopropylthiogalactoside at 37 °C for 16 h, harvested by centrifugation, lysed by sonication and the soluble proteins removed by centrifugation at 40,000g. The cell pellet was sequentially extracted with 1 M NaCl, 1% Triton X-100 and 1.75 M guanidinium-HCl. The SC1N β AR fusion protein was extracted from the cells with 3.25 M guanidinium-HCl, dialysed against phosphate-buffered saline and used directly as an immunogen. Approximately 100 mg of fusion protein of 90% purity was obtained from 1 litre of starting culture. Antibodies were detected in serum from injected rabbits by incubation of the serum with 125 I-CYP-labelled soluble β AR in 10 mM Tris-HCl, 0.1 M NaCl, 0.1% digitonin, 0.5% bovine serum albumin (BSA) pH 7.4. After 2 h at 25 °C, the antibody was precipitated by addition of either $(\text{NH}_4)_2\text{SO}_4$ to 50% or *Staphylococcus aureus* protein A, followed by incubation in ice for 30 min. The precipitated protein was collected by centrifugation, and the radioactivity contained in the antibody pellet measured. For the peptide blocking experiment, the peptide YAKRQLQKIDKSEGR was synthesized²⁹ using a SAM II peptide synthesizer (Biosearch), and purified on a Whatman C-18 Magnum column in a H_2O /acetonitrile gradient containing 0.2% TFA. The resulting product was judged to be pure by amino-acid sequencing and mass spectral analysis. Increasing concentrations of this peptide were added to a 1:100 dilution of anti-peptide 7 antiserum and incubated for 2 h at 23 °C. The treated antiserum was then mixed with 125 I-CYP-labelled β AR and assayed as above.



nizing the amino-acid sequence of peptide 7, a chemically synthesized peptide was used as a specific inhibitor of antibody-receptor interactions. At concentrations $\geq 100 \mu\text{M}$, this synthetic peptide completely prevented the immunoprecipitation of 125 I-CYP-labelled β AR by the antibody (Fig. 2*b*). An unrelated peptide, atrial natriuretic factor⁹, had no effect on the immunoprecipitation.

The specificity of the antibody for the β AR was demonstrated further by protein immunoblotting. As shown in Fig. 2*c* (lane 1), the antibody reacted specifically with pure hamster lung β AR; a single protein of the same relative molecular mass (64K) was also observed in human A431 and hamster melanoma RPMI 1846 cells (Fig. 2*c*, lanes 2, 3), both of which were found to contain β AR on the basis of 125 I-CYP binding¹⁰ (data not shown). This specific immunoreactive band was not observed on prior treatment of the antibody with the synthetic peptide and was not present when normal rabbit serum was substituted for the anti-peptide 7 antibody (data not shown).

To facilitate cloning of the β AR gene, oligonucleotides complementary to the DNA encoding the amino-acid sequence of peptide 7 were synthesized for use as hybridization probes (see Fig. 3 legend). In hybridization experiments performed at high

stringency on blots of hamster genomic DNA, a single hybridizing band of 5.2 kilobases (kb) was observed in *EcoRI* digests and a band of 1.3 kb was observed in *HindIII* digests (data not shown). When a complete hamster genomic library was screened under the same conditions, five clones were isolated. Restriction analysis of the phage DNA revealed that all these clones contained a 1.3-kb *HindIII* and a 5.2-kb *EcoRI* fragment which hybridized to the probes (data not shown). Mapping of the phage DNA indicated that these clones overlap to give a total of 30 kb of contiguous genomic DNA. Figure 3 shows the restriction map of the genomic DNA containing the β AR-related sequences. Sequencing of the 1.3-kb *HindIII* fragment revealed a continuous open reading frame encoding 435 amino acids; the sequences of all the CNBr peptides shown in Fig. 1 were contained within this putative polypeptide.

Using the 1.3-kb *HindIII* gene fragment as a probe, seven clones were obtained from an unamplified hamster cDNA library (2×10^6 recombinants). Two of these cDNAs hybridized to oligonucleotide probes specific for the N-terminal, middle and C-terminal portions of the β AR gene. The nucleotide sequence of these two cDNA clones (Fig. 4) extends from 210 nucleotides (nt) 5' to the open reading frame encoding the β AR

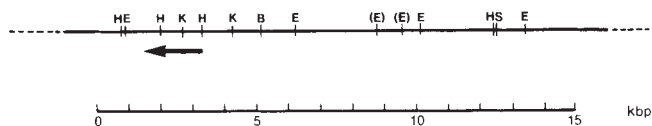


Fig. 3 Restriction map of the hamster β AR gene. A portion of the hamster DNA is shown. The 1.3-kb *Hind*III fragment which hybridizes to oligonucleotides specific for peptide 7, is underlined with an arrow indicating the direction of transcription of the β AR gene. The restriction enzyme sites are: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; S, *Sal*I. Those sites shown in parentheses have not been unequivocally ordered.

Methods. All restriction enzymes, *E. coli* DNA polymerase I, T_4 DNA ligase and T_4 polynucleotide kinase were purchased from New England Biolabs. Radiolabelled nucleotides were purchased from Amersham. λ EMBL3A phage arms³⁰ and λ *in vitro* DNA packaging extracts were purchased from Vector Cloning Systems. Standard recombinant DNA and microbiological procedures were used throughout²⁸. Genomic libraries were constructed using high- M_r genomic DNA isolated from hamster lung cells in the vector λ EMBL3 (refs 6, 30). Probes for peptide 7 coding sequences, oligonucleotides ON225(5' pTCCACCTGGCCAGGTTGGG-AGAGTGGAACTGCCCTCAGACTTGTCGAT) and ON229(5' pAGGCAGCTGCAGAAGATCGACAAGTCTGAG) and ON168 (5' pTTCCAGGTGGCCCAAGCGGCAGCTGCAGAAGATCGACAA) and ON169(5' pATGGTCTTTGTCTACTCCCGGGTCT-TCCAGGTGGCCAA), were synthesized on an Applied Biosystems Model 3A DNA synthesizer. The oligonucleotides were labelled by either a fill-in reaction (ON225, ON229) using Klenow DNA polymerase and all four [α - 32 P]dNTPs³¹ or by phosphorylation using T_4 polynucleotide kinase and [γ - 32 P]ATP²⁸. Phage libraries were screened by the method of Benton and Davis³² using the hybridization conditions of Ullrich *et al.*³³. DNA was isolated from CsCl-banded phage as described elsewhere²⁸. For restriction analysis, DNA was digested with the appropriate enzyme and electrophoresed on 0.8% agarose gels. DNA was transferred to nitrocellulose by the procedure of Southern³⁴ and hybridized as above.

Fig. 4 (Right) Nucleotide and deduced amino-acid sequence of the β AR cDNA. The nucleotides are numbered on the right-hand side of each line beginning with the first nucleotide of the most 5' cDNA clone. The translated amino-acid sequence is shown beneath the corresponding nucleotide sequence and is numbered to the left of each line. Underlined amino acids represent the CNBr peptides whose sequences are given in Fig. 1. All predicted amino acids agree with those determined by peptide sequencing, with the single exception of a cysteine for serine substitution in peptide 7a. All derived peptide sequences are preceded by a methionine, consistent with CNBr cleavage. The underlined nucleotides preceding the β AR sequences denote the first methionine codon and an in-frame termination codon. The boxed nucleotides at the 3' end of the sequence represent the polyadenylation signal. Postulated glycosylation sites (Asn-X-Ser/Thr) are indicated by asterisks. Putative protein kinase A phosphorylation sites are boxed.

Methods. Reverse transcriptase was purchased from Seikagaku-America. RNase H was from Pharmacia and λ gt10 arms were from Vector Cloning. Total cellular RNA was isolated from growing cultures of DDT1-MF2 cells by the guanidinium isothiocyanate-CsCl method⁴⁷. Poly(A)⁺ RNA was purified by chromatography on oligo(dT)-cellulose. Double-stranded cDNA was synthesized by oligo(dT)-primed reverse transcription of the poly(A)⁺ RNA, followed by treatment with *E. coli* DNA polymerase and RNase⁴⁸. The ends of the cDNA were blunted with T_4 DNA polymerase²⁸. Following protection of the *Eco*RI sites with *Eco*RI methylase, *Eco*RI linkers were added²⁸. *Eco*RI-digested cDNA was size-fractionated by agarose gel electrophoresis to obtain cDNAs between 2 and 7 kb long. The cDNA was ligated to the vector λ gt10 (ref. 49) and packaged *in vitro*. The resulting library was screened unamplified as described in Fig. 3 legend. The 1.3-kb *Hind*III fragment was labelled using [α - 32 P]dCTP by nick-translation for use as a hybridization probe²⁸. The cDNA inserts contained in the positive phage were subcloned into pUC13 or M13 mp19 for DNA sequence analysis³⁵⁻³⁷. Both strands of the clones were sequenced with no discrepancies.

70	CAGCGTTCAA	GCTGCTGTTA	GCAGGCACCG	CGAGCCCCGG	GCACCCACG	AGCTGAGTGT	GCAGGACGGC
140	CCCCCAGCAG	AGCCACCTAC	AGCCGCTGAA	TGAAGCTTCC	AGGAGTCTGC	CTCCGCCCGC	CTCGGCCCGC
210	TCGGAGGTGC	ACCCGCTGAG	AGCGCCAGGG	CACCAGAAAG	CCGGTGCCTG	CACCTGCTCG	TCTGCCACGG
264	ATG	GGG	CCA	CCC	GGG	AAC	GAC
CAT	Gly	Pro	Pro	Gly	Asn	Asp	Ser
318	GTG	CCA	GAC	CAC	GAT	GTC	ACT
ATG	Val	Pro	Asp	His	Asp	Val	Thr
372	GCC	ATC	CTT	ATG	TCG	GTT	ATC
GTC	Ala	Ile	Leu	MET	Ser	Val	Ile
426	ATC	ACA	GCC	ATT	GCC	AAG	TTC
ATA	Ile	Thr	Ala	Ile	Ala	Lys	Phe
480	ACC	TCC	TTG	GCG	TGT	GCT	GAT
GGG	Thr	Ser	Leu	Ala	Cys	Val	Thr
534	GCC	AGT	CAC	ATC	CTT	ATG	AAA
TTC	Ala	Ser	His	Ile	Leu	MET	Lys
588	TGG	ACT	TCC	ATT	GAT	GTG	TTA
GTC	Thr	Ser	Thr	Ser	Ile	Asp	Val
642	ATA	GCA	GTG	GAT	CGC	TAC	ATT
GTC	Ile	Ala	Val	Asp	Arg	Tyr	Ile
696	CTG	ACC	AAG	AAT	AAG	GCC	CGA
GGC	Leu	Thr	Lys	Asn	Lys	Ala	Arg
750	CTT	ACC	TCC	TTC	TTG	CCC	ATT
Lys	Thr	Ser	Phe	Leu	Pro	Ile	Gln
804	GCC	ATC	GAC	TGC	TAT	CAC	AAG
Ala	Ile	Asp	Cys	Tyr	His	Lys	Glu
858	TAC	GCC	ATT	GCT	TCC	TCC	ATT
Val	Tyr	Ala	Ile	Ala	Ser	Ser	Ile
912	TTT	GTC	TAT	TCC	AGG	GTC	TTC
ASP	Phe	Val	Tyr	Ser	Arg	Val	Phe
966	AAA	TCT	GAG	GGA	AGA	TTC	CAC
GGG	Lys	Ser	Glu	Gly	Arg	Phe	His
1020	CGG	AGT	GGG	CAC	GTC	CTC	CGA
AAA	Arg	Ser	Gly	His	Gly	Leu	Arg
1074	GCC	CTC	AAG	ACT	TTA	GGC	ATC
CCC	Ala	Leu	Lys	Thr	Leu	Gly	Ile
1128	TTT	TTT	ATT	GTC	AAC	ATC	GTG
Glu	Phe	Phe	Ile	Val	Asn	Ile	Val
1182	GTT	TAC	ATC	CTC	CTT	AAC	TGG
CTC	Val	Tyr	Ile	Leu	Leu	Asn	Trp
1236	ATC	TAC	TGT	CGG	AGT	CCA	GAT
CTC	Ile	Tyr	Cys	Arg	Ser	Pro	Asp
1290	CGC	AGG	TCT	TCT	TCA	AAA	GCC
Gly	Arg	Ser	Ser	Ser	Lys	Ala	Tyr
1344	AAA	ACA	GAC	TAC	ATG	GGG	GAG
GAA	Lys	Thr	Asp	Tyr	MET	Gly	Glu
1398	AGT	GAA	CGG	CTG	TGT	GAG	CCC
CAA	Ser	Glu	Arg	Leu	Cys	Glu	Asp
1452	GGT	ACT	GTG	CCT	AGC	CTT	AGC
AAT	Gly	Thr	Val	Pro	Ser	Leu	Asp
1517	GAC	TCA	CCG	CTG	TAA	TGCAGGCTTT	CTGCTTTTAA
1587	CTATTTAAC	TGAGTGAAT	AACCTTAGAA	TAAACTGTA	TAGAGATTG	CAGAAGGGGA	GCATCCTTCT
1657	GCCCTTTT	ATTTATTT	TTTAAGCCG	AAAAATAGAG	AGGGAGAGAA	ACTGTACTTG	AGTCTTGT
1727	TGTTCTGT	GCAATTCAGT	TCCTCTTTCG	GTGGAACCTA	AAAGTTCTG	TCTGAAGTAT	GTGGGTCT
1797	AGAGGAGTGT	CTGTATGTTT	AGATGATTTT	CCATGCATCT	ACCTCACTCG	TCAAGTGTTA	GGGGATACCG
1867	TGCTAGTAAT	TGTACCTGA	AGGAAATTTT	CCTTCTGTA	CCCTTACACT	TGTCAATCCT	GTGCTTGGA
1937	CCTTCTGCT	GTGAATATAT	ACTCTCTCCC	GCTCCACTTA	TTTGCTCAAA	TGGAGTGTGT	AGACAGGGAT
2007	CTTGAGGCAG	AGCTTCAGTT	GGTTTTTTTT	TTTTTTTGA	GCAAAGTCTA	AAGTTTACAG	TATATAAAT
2076	GTTTGACCAC	GAATAAAAAA					

peptides to 560 nt 3' to the termination codon. A hexanucleotide AATAAA occurs near the 3' end, followed by a poly(A) stretch. The nucleotide sequence of the genomic clones is identical to that of the cDNA clones up to but not including the poly(A) tail. These data demonstrate an absence of introns within the coding and 3'-untranslated regions of the β AR gene but the possibility of introns in the untranslated region 5' to the sequence isolated cannot be ruled out. While the lack of introns is unusual, it is not unprecedented in that sea urchin histone genes³⁸ and mammalian α - and β -interferon genes^{39,40} are also uninterrupted. Previous studies with simian virus 40 and β -globin indicate that introns are necessary for efficient expression of those genes^{41,42}, so the lack of introns within the hamster β AR gene may account in part for its low level of expression.

Whereas most eukaryotic genes are translated using the first AUG encountered in the messenger RNA⁴³, the open reading frame (ORF) encoding the β AR peptides begins at the second AUG. The first AUG is followed by a termination codon after only 19 amino acids. This open reading frame is in-frame with the AUG codon beginning the β AR polypeptide. A similar situation has recently been reported for the oestrogen receptor, where the reading frame of that receptor is preceded by a short 20-amino-acid ORF⁴⁴. Note that the DNA sequence around the AUG codon for the second ORF agrees well with the consensus eukaryotic translation initiation sequence⁴³ (CAGCGAUGG compared with CCGCCAUGG), whereas the sequence surrounding the first AUG does not.

Translation beginning with the second AUG as the initiation site would produce a polypeptide of 418 amino acids, with a M_r (46K) in close agreement with the apparent M_r of the deglycosylated β AR (49K)¹¹. The protein sequence immediately following the initiator methionine residue is identical to the sequence determined for peptide 2. Apparently, like several other integral membrane proteins—for example, bovine opsin⁴⁵—the β AR does not contain a cleavable signal sequence and may use internal signals for the insertion of the protein into the membrane. The receptor has been shown to have two sites of *N*-linked glycosylation (R.J.L. *et al.*, in preparation). Both sites are present in peptide 2, consistent with the results of peptide sequencing. The presence of two consensus protein kinase A and C sites⁴⁶ within the coding region (Fig. 4) agrees with *in vitro* phosphorylation results²⁰.

Hydropathicity profiles of the predicted β AR amino-acid sequence were produced using the analyses of Hopp and Woods¹² and of Kyte and Doolittle¹³, with similar results. As shown in Fig. 5a, the β AR sequence should encode a largely hydrophobic polypeptide, with the N-terminal region of the receptor being predominantly hydrophobic and the C-terminal region of the molecule being hydrophilic. The β AR hydropathicity profile is remarkably similar to that of the rhodopsins (Fig. 5b), of which bacteriorhodopsin⁷ is known to contain seven membrane-spanning helices. Not only does a similar pattern of repeating hydrophobic sequences 20–25 residues long occur in the predicted β AR sequence, but also the amino-acid composition of these postulated helices is similar to that of the rhodopsins, having a high proportion of proline and aromatic amino acids. The exact number of transmembrane regions remains to be determined. Amino-acid homology was apparent when the sequences corresponding to the postulated helices V, VI and VII of bovine opsin¹⁴, which comprise the retinal binding site¹⁵, were aligned with those for the analogous regions of β AR (Fig. 5c).

The sequence homology between β AR and rhodopsin parallels similarities in their function: both rhodopsin and β AR are involved in signal transduction mechanisms that involve interaction with the guanine nucleotide regulatory proteins transducin¹⁶ and G_s (ref. 17), respectively. Moreover, it seems that phosphorylation has an important role in the regulation of both rhodopsin and β AR^{18,19}. Rhodopsin is multiply phosphorylated at its C-terminus, which contains seven serine

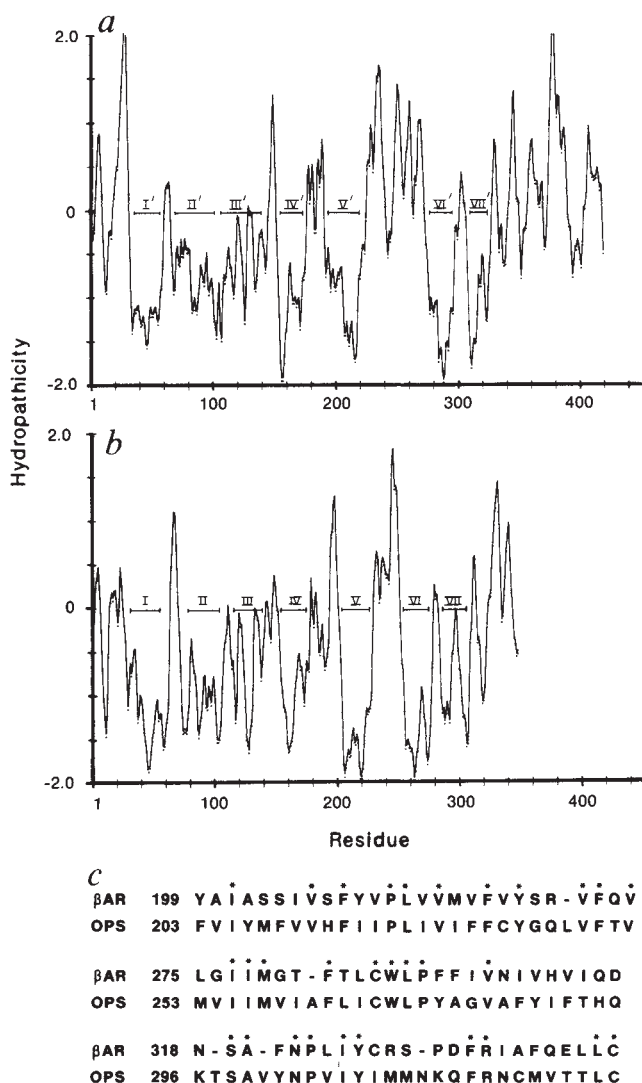


Fig. 5 Structure and sequence homology for hamster β AR and bovine opsin. Hydrophaticity profiles are shown for hamster β AR (a) and bovine opsin¹⁴ (b). Values were calculated by the method of Hopp and Woods¹⁴. Hydrophobicity increases with decreasing values. Horizontal lines indicate hydrophobic peptide regions 20–25 residues long. The putative transmembrane helices of bovine opsin are numbered as in ref. 14. We have designated the proposed transmembrane helices of β AR based on the analogous hydrophaticity patterns of β AR and opsin. c, Amino-acid sequence homologies between the hamster β AR and bovine opsin (OPS). Positions of amino-acid identities are designated by asterisks. The underlined lysine residue (296) of bovine rhodopsin is involved in a Schiff base with retinal¹⁵.

and threonine residues¹⁸. The C-terminus of the postulated sequence for β AR contains several serine and threonine residues that could serve as sites for phosphorylation¹⁹.

The similarity in both the amino-acid sequence and proposed structure for the retinal binding site of rhodopsin and the analogous region in β AR suggests a specific mode of action for the ligands that modulate β AR function. We propose that these compounds interact with the β AR in a manner similar to that by which retinal interacts with opsin²¹; that is, they intercalate among the hydrophobic transmembrane helices and thereby determine whether the receptor is in its active or inactive conformation. According to this model, agonists and antagonists would mimic the action of photoactivated and native retinal. The generality of this hypothesis will be tested once the genes for similar membrane-bound receptors, such as those for leuko-

trienes, prostaglandins, dopamine and histamine, have been isolated. Our proposed model for the structure of β AR and its interaction with pharmacologically important ligands should, together with the biochemical and genetic studies now possible, provide a rational basis for a new approach to the development of more selective drugs.

We thank Lenora Davis, Jill D'Alonzo, Ester Hou, Tong Sun Kobilka, John Rodkey, Claudia Staniszwski, H. Vincent Strout Jr and Barbara Zemcik for technical assistance, Jeanne White and Lynn Tilley for manuscript preparation, and Drs Eugene Cordes, Ronald Ellis, Russell Kaufman, Paul Keller, Pete Kniskern, Richard Lebovitz, Mark Riemen and Michael Rosenblatt for helpful discussions. We particularly thank Dr Edward Scolnick for continued support and direction.

Received 10 March; accepted 11 April 1986.

1. Lefkowitz, R. J., Stadel, J. M. & Caron, M. G. *A. Rev. Biochem.* **52**, 159-186 (1983).
2. Shorr, R. G., Lefkowitz, R. J. & Caron, M. G. *J. biol. Chem.* **256**, 5820-5826 (1981).
3. Shorr, R. G., Strohsacker, M. W., Lavin, T. N., Lefkowitz, R. J. & Caron, M. G. *J. biol. Chem.* **257**, 12341-12350 (1982).
4. Homey, C. J., Rockson, S. G., Countaway, J. & Egan, D. A. *Biochemistry* **22**, 660-668 (1983).
5. Benovic, J. L., Shorr, R. G., Caron, M. G. & Lefkowitz, R. J. *Biochemistry* **23**, 4510-4518 (1984).
6. Cerione, R. A. *et al. J. biol. Chem.* **259**, 9979-9982 (1984).
7. Engelman, D. M., Goldman, A. & Steitz, T. A. *Meth. Enzym.* **88**, 81-88 (1982).
8. Temes, G. L., Gibbs, J. B., D'Alonzo, J. S., Sigal, I. S. & Scolnick, E. M. *Nature* **313**, 700-703 (1985).
9. Seidah, N. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 2640-2644 (1984).
10. Delavier-Klutchko, C., Hocheke, J. & Strasberg, A. D. *FEBS Lett.* **169**, 151-155 (1984).
11. Stiles, G. L., Benovic, J. L., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **259**, 8655-8663 (1984).
12. Hopp, T. P. & Woods, K. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3824-3828 (1981).
13. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105-132 (1982).
14. Nathans, J. & Hogness, D. S. *Cell* **34**, 807-814 (1983).
15. Hubbell, W. L. & Bownds, M. D. *A. Rev. Neurosci.* **2**, 17-34 (1979).

16. Stryer, L. *A. Rev. Neurosci.* **9**, 87-119 (1986).
17. Gilman, A. G. *Cell* **36**, 577-579 (1984).
18. Hargrave, P. A. *Prog. Retinal Res.* **1**, 1-51 (1982).
19. Benovic, J. L., Strasser, R. H., Caron, M. G. & Lefkowitz, R. J. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
20. Benovic, J. L. *et al. J. biol. Chem.* **260**, 7094-7101 (1985).
21. Thomas, D. D. & Stryer, L. *J. molec. Biol.* **154**, 145-157 (1982).
22. Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **251**, 2374-2384 (1976).
23. Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. J. *J. biol. Chem.* **256**, 7790-7797 (1981).
24. Spiess, J., Rivier, J. E., Rodkey, J. A., Bennett, C. D. & Vale, W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2974-2978 (1979).
25. Laemmli, U. K. *Nature* **117**, 680-685 (1970).
26. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4356 (1979).
27. Johnson, D. A. & Elder, J. H. *J. exp. Med.* **159**, 1751-1756 (1983).
28. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
29. Merrifield, R. B. *J. Am. chem. Soc.* **85**, 2149-2154 (1963).
30. Frischauf, A. M., Lehrach, H., Paustka, A. & Murray, N. *J. molec. Biol.* **170**, 827-842 (1983).
31. Lauffer, L. *et al. Nature* **318**, 334-338 (1985).
32. Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
33. Ullrich, A., Berman, C. H., Dull, T. J., Gray, A. & Lee, J. M. *EMBO J.* **3**, 361-364 (1984).
34. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
35. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-565 (1977).
36. Vieira, J. & Messing, J. *Gene* **19**, 259-268 (1982).
37. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
38. Schaffner, W. *et al. Cell* **14**, 655-671 (1978).
39. Nagata, S., Mantei, N. & Weissman, C. *Nature* **287**, 401-408 (1980).
40. Lawn, R. M. *et al. Nucleic Acids Res.* **9**, 1045-1052 (1981).
41. Volckaert, G., Feunteun, J., Crawford, L. V., Berg, P. & Fiers, W. J. *J. Virol.* **30**, 674-682 (1979).
42. Hamer, D. H. & Leder, P. *Cell* **18**, 1299-1302 (1979).
43. Kozak, M. *Nucleic Acids Res.* **12**, 857-872 (1984).
44. Green, S. *et al. Nature* **320**, 134-139 (1986).
45. Friedlander, M. & Blobel, G. *Nature* **318**, 338-343 (1985).
46. Feramisco, J. L., Glass, D. B. & Krebs, E. G. *J. biol. Chem.* **255**, 4240-4245 (1980).
47. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
48. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
49. Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning a Practical Approach* Vol. 1, 49-78 (IRL, Oxford, 1985).

In vitro osteoclast generation from different bone marrow fractions, including a highly enriched haematopoietic stem cell population

Ben A. A. Scheven*, Jan W. M. Visser† & Peter J. Nijweide*

* Laboratory of Cell Biology and Histology, University of Leiden, Rijnsburgerweg 10, 2333 AA Leiden, The Netherlands
† Radiobiological Institute TNO, Lange Kleiweg 151, 2288 GJ Rijswijk, The Netherlands

It is well established that the osteoclast is formed by fusion of post-mitotic, mononuclear precursors¹ derived from circulating progenitor cells². However, the precise haematopoietic origin of the osteoclast is unknown. We have investigated this here by fractionating mouse bone marrow and isolating haematopoietic stem cells using a three-step method combining equilibrium density centrifugation and two fluorescence-activated cell sortings (FACS)³, and have tested the ability of each bone marrow fraction, including highly purified haematopoietic stem cells, to generate osteoclasts during co-culture with preosteoclast-free embryonic long bones^{4,5}. The osteoclast-forming capacity was found to increase with increasing stem cell purity. On the other hand, the culture time needed for osteoclast formation also increased with purification, suggesting the presence of progressively more immature progenitor cells. The pluripotent haematopoietic stem cell fractions with the highest purity needed preincubation with a stem cell-activating factor (interleukin-3) to activate the predominantly quiescent stem cells *in vitro*.

Embryonic long bones stripped of the endogenous osteoclast precursor pool (the periosteum) were used to induce the development of osteoclasts from different bone marrow cell populations. When stripped long bones are co-cultured with cell populations containing osteoclast progenitors, osteoclasts

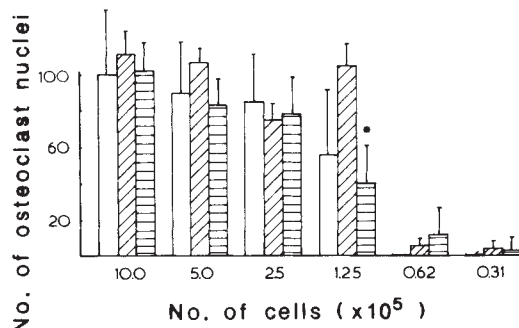


Fig. 1 Osteoclast formation in relation to the number of bone marrow cells (per plasma clot per bone) after 8 days of co-culture. Means $\pm$ s.d. of the number of osteoclast nuclei in five axial sections of four to six cultures are shown. □, Unfractionated bone marrow cells; ▨, low-density bone marrow cells; ■, high-density bone marrow cells. Statistically significant ($P < 0.05$) compared with low-density cells.

develop which thereby give rise to the formation of a marrow cavity. Culture of stripped bones alone does not lead to appearance of osteoclasts^{4,5}. Total bone marrow cell population and both high-density ($1.100 < \rho < 1.078 \text{ g cm}^{-3}$) and low-density ($\rho < 1.078 \text{ g cm}^{-3}$) bone marrow fractions, obtained by equilibrium density centrifugation on a discontinuous metrizamide gradient, gave rise to osteoclast formation (Table 1). Autoradiography using continuous ³H-thymidine labelling indicated that the osteoclasts were derived from progenitor cells which had proliferated during co-culture. The osteoclasts exhibited tartrate-resistant acid phosphatase activity, which is specific for these cells. Some mononuclear cells and an occasional multinucleated cell present in the plasma clot surrounding the bones also displayed this osteoclastic enzyme activity. Mature granulocytes accumulated in the developing marrow cavity following the invasion of osteoclasts. With time, other cell types such as megakaryocytes, macrophages and undifferentiated cells were observed in the expanded cavity.

Dilution experiments were performed to assess the relationship between the number of cells added to the bones and the

Table 1 Osteoclast formation from different bone marrow fractions

	No. of osteoclast nuclei per bone					No. of cells added per bone ($\times 10^5$)	No. of day-12 CFU-S per 10^5 cells	Culture time (days)
	0	1-5	6-20	>20	(n)			
Unfractionated bone marrow cells				26	26	2.5-10.0	51 ± 5	7-8
Low-density cells				22	22	2.5-10.0	ND	7-8
High-density cells			5	17	22	2.5-10.0	ND	7-8
Total WGA ⁺ cells	2			12	14	2.0-3.0	ND	7
Total WGA ⁻ cells	14	1			15	2.0-3.0	ND	7
WGA ⁺ low-density cells		1		4	5	0.40-0.50	ND	9-11
WGA ⁻ low-density cells	8	1			9	0.40-0.50	ND	9-11
WGA ⁺ and H-2K ⁺ low-density cells (-SAF)	7	1			8	0.0125	690 ± 840	14-17
WGA ⁺ and H-2K ⁺ low-density cells (+SAF)	1		1	10	12	0.0250	$2,900 \pm 510$	2+8

Metatarsal bones of 17-day-old mouse embryos were dissected and stripped of surrounding periosteum using the collagenase and mechanical treatment of Burger *et al.*⁴. The stripped cartilaginous bone rudiments consisted of a central calcified area (without (pre)osteoclasts and marrow cavity) with a thin bone collar and two uncalcified epiphyses. The bone marrow cell preparation and separation procedures were performed as described elsewhere³. The cells were first separated by density using a discontinuous metrizamide gradient and simultaneously labelled with WGA and fluorescein isothiocyanate (FITC), and then collected at the interfaces between the 1.100 and 1.078 g cm⁻³ solutions (high-density fraction) and the 1.078 and 1.055 g cm⁻³ solutions (low-density fraction). Cells from unfractionated bone marrow and the low-density cells were analysed for WGA-FITC binding by a modified fluorescence activating cell sorter (FACS II, Becton Dickinson). All WGA-FITC-positive cells with medium forward light scatter and low perpendicular light scatter were sorted. WGA-FITC was removed from the sorted cells by treatment with 0.2 M *N*-acetyl-D-glucosamine for 15 min at 37 °C. The cells were then labelled with biotinylated anti-H-2K followed by avidin-FITC. The most brightly fluorescent cells were selected by FACS. The low-density cells comprised ~15% of all nucleated bone marrow cells. About 30-40% of the total bone marrow was WGA⁺, compared with about 4-6% of the low-density cells. The H-2K⁺ fraction comprised ~0.3% of all nucleated bone marrow cells. SAF (10 U) was used for preincubation of the isolated H-2K⁺ cells for 2 days before co-culture with stripped embryonic bones. 1 U of SAF is the amount needed for 50% proliferation of the interleukin-3-dependent cell line DA-1. The isolated cell suspensions of known cell number were diluted with 10% rat serum and 10% chicken embryo extract in HEPES-buffered Hanks' balanced salt solution and centrifuged to collect the cells. Then 10 μ l of cock plasma was added to the cells to make a weak plasma clot. The clot was divided into equal parts of ~1 mm³ and placed along both sides of the stripped bones. Co-culture was performed on a semi-solid medium, composed of 60% Eagle's α -minimal essential medium, 10% rat serum, 10% chicken embryo extract and 20% cock plasma, at 37 °C in a humidified 5% CO₂ atmosphere. The numbers of osteoclast nuclei in the marrow cavity were counted in five axial histological sections of bone rudiments. Recipient mice for the CFU-S assay¹⁰ received 9.5 Gy total body irradiation (¹³⁷Cs, 0.85 Gy min⁻¹) 4-6 h before transplantation. Spleen colonies were counted on day 12. ND, not determined.

number of osteoclasts formed, and to determine the minimum number of cells needed for osteoclast formation (Fig. 1). At high cell concentrations the number of osteoclast nuclei reached about the same plateau level for all three bone marrow fractions. However, at a cell dose of 1.25×10^5 cells per clot, the low-density cells showed a significantly greater potency for osteoclast formation than the unfractionated and high-density bone marrow cells. The experiments were repeated twice at this critical cell dose, resulting in 89.3 ± 25.0 and 17.3 ± 11.2 osteoclast nuclei for the low- and high-density cells, respectively (means $\pm$ s.d. of five axial sections, $P < 0.001$, $n = 6$). Lower cell doses led to an abrupt decline in the number of osteoclasts developing during culture from these three fractions.

Examination of the co-cultures using a diavert microscope revealed that resorption of the calcified centre generally started earlier when the bones had been co-cultured with the high-density bone marrow cells than with the low-density cells. To quantitate this, osteoclastic mineral resorption was determined by measuring the release of ⁴⁵Ca from prelabelled bones during co-culture. These experiments were carried out using high cell doses so as to exclude possible cell concentration effects (Fig. 1). Figure 2 convincingly demonstrates that the high-density cells indeed gave rise to an initial, significantly higher release of ⁴⁵Ca. However, from day 6 onwards the relative ⁴⁵Ca release was equal in all groups.

Our results indicate that the different bone marrow density fractions contain subpopulations of osteoclast progenitor cells which differ in maturation stage. The low-density fraction apparently contains more immature progenitor cells which have

a greater proliferation capacity (Fig. 1) but need a longer time to differentiate into mature, fusible preosteoclasts (Fig. 2). This is consistent with the finding that the primitive progenitors, type 1 granulocyte-macrophage colony-forming units (GM-CFU-1) and spleen colony-forming units (CFU-S), can be separated from the more mature haematopoietic cells by their low buoyant density⁶⁻⁹. At increasing buoyant densities the proliferation potency of the culture colony-forming units (CFU-C) decreases, whereas the maturation rate increases^{7,8}. The earlier ⁴⁵Ca release induced by the high-density cells correlates with the earlier maturation of agar colonies from GM-CFU-2 and -3, while the late osteoclast formation from the low-density cells may be compared to colony formation from GM-CFU-1.

Sorting on the basis of labelling of unfractionated bone marrow cells by wheat germ agglutinin (WGA) resulted in a distinct separation between an osteoclast-forming (WGA⁺) and a non-osteoclast-forming (WGA⁻) fraction (Table 1). The total WGA⁺ cell population comprised cells belonging to the granulocyte series, including the mature forms, immature cells of the monocyte-macrophage lineage and early haematopoietic progenitor and stem cells. The WGA⁻ cell population, on the other hand, consisted of erythrocytes, lymphocytes, megakaryocytes and their immediate precursors, and mature macrophages and osteoclasts.

The WGA⁺ fraction of the low-density cells consisted mainly of early haematopoietic progenitor and stem cells with 60-fold enrichment of CFU-S⁸. Only the WGA⁺ low-density cells were able to induce osteoclast development (Table 1). This, however, took a longer co-culture period (8-10 days) compared with the

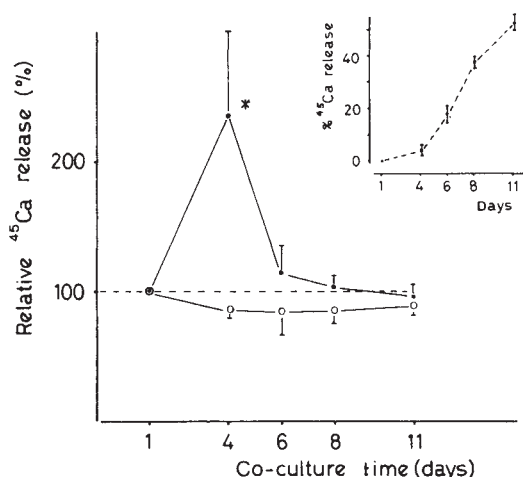


Fig. 2 Osteoclastic mineral resorption during co-culture with high (●) or low (○)-density bone marrow cells. At each time point the release of ⁴⁵Ca from prelabelled metatarsal bones is expressed as a percentage of the average ⁴⁵Ca release obtained during co-culture with unfractionated cells. Inset, cumulative release of ⁴⁵Ca during co-culture with unfractionated bone marrow cells shown as a percentage of total incorporated ⁴⁵Ca per bone.

Methods. Pregnant mice were injected with 50 μ Ci ⁴⁵Ca on day 16 of gestation. The next day the ⁴⁵Ca-labelled metatarsal bones were dissected, stripped, explanted on a semi-solid medium and co-cultured with clots containing unfractionated, low-density or high-density bone marrow cells (see Table 1 legend). Medium was changed after 1, 4, 6 and 8 days. After a total culture period of 11 days the co-cultures were fixed in a mixture of 5% glacial acetic acid, 85% ethanol and 10% formaldehyde for 24 h, then decalcified in a 5% formic acid/5% formaldehyde solution. The semi-solid media were digested overnight with 0.1% trypsin and the digested culture media, fixation and decalcification fluids analysed for ⁴⁵Ca radioactivity on a Beckman LS 1800 liquid scintillation counter. The release of ⁴⁵Ca during the first culture day, which represents physicochemical exchange, was discarded. The values shown are means $\pm$ s.e.m. of seven or eight co-cultures. * Statistically significant ($P < 0.05$) compared with low-density cells.

cell populations described above. This phenomenon is probably due to the lower number of cells per bone and the occurrence of predominantly very immature cells in this fraction which needed more time to mature into preosteoclasts.

H-2K sorting of the WGA⁺ low-density fraction produced a highly enriched pluripotent haematopoietic stem cell population. Intravenous injection of 100 sorted cells per lethally irradiated mouse resulted in 6.9–9.3 colonies per spleen 12 days later (day-12 CFU-S; refs 3, 10). As the seeding efficiency of transplanted haematopoietic stem cells to the spleen is between 0.05 and 0.10 (refs 3, 11, 12), this indicates that at least 69–93% of the sorted cells were pluripotent stem cells.

The osteoclast-forming capacity of these sorted cells, however, was minimal (Table 1), perhaps because of the lower number of cells present per clot and/or the absence of suitable factors for activating the mainly quiescent stem cells. A stem cell-activating factor (SAF=interleukin-3)¹³ was therefore used to stimulate the sorted stem cells. Preincubation of cells with 10 U SAF for 2 days resulted in a doubling of the total number of cells, concomitantly with maintenance of day-8 CFU-S and a slight decrease in the number of day-12 CFU-S per culture. Subsequent co-culture with stripped metatarsal bones led to a burst of haematopoietic cell growth (granulocytes and macrophages) and abundant osteoclast formation. This indicates either that SAF triggers the pluripotent haematopoietic stem cells into a state in which they are responsive to the growth factors in the plasma clot, or that the pluripotent stem cells, because of SAF, differentiate into committed progenitors which are then stimulated by the clot. The plasma clot itself apparently does not contain SAF. When, in another experiment, we added

recombinant SAF (10 U ml⁻¹) to a co-culture of purified stem cells and stripped metatarsal bones on day 0 and removed the SAF on day 2 by washing, we observed extensive haematopoietic cell growth and abundant osteoclast formation similar to that in the above experiment. Without SAF no osteoclast formation was observed in this experiment either.

Our results with purified stem cells provide direct, although not yet conclusive, evidence for the hypothesis that osteoclasts are derived from the pluripotent haematopoietic stem cell^{14–17}. In addition, the data indicate that SAF is needed to initiate the proliferation and differentiation of the pluripotent stem cell. The evidence presented cannot exclude the possibility that the osteoclast progenitor co-purifies with the CFU-S population, although we consider this highly unlikely. It is not very probable that a distinct osteoclast stem cell would behave exactly like the haematopoietic stem cell during the isolation procedure, causing a >100-fold enrichment of the latter cell, for it would mean the possession of similar cytochemical (WGA, H-2K) and physical (density, light-scattering) properties, and the ability to react in a similar manner to SAF.

The first part of the pathway along which osteoclasts differentiate may very well coincide with the monocyte differentiation lineage, as granulocyte and macrophage precursors are also found among the WGA⁺ bone marrow cells¹⁸. However, as we have shown earlier¹⁹, the direct precursors of the osteoclast are not monocytes. The formation of osteoclasts from isolated haematopoietic stem cells in the *in vitro* model used here should facilitate the elucidation of the osteoclast differentiation lineage and its relation to other blood cell lineages.

We thank Dr L. Dorssers for supply of (recombinant) SAF, Mrs A. M. Wassenaar, E. W. M. Kawilarang-de Haas and M. Hogeweg-Platenburg for technical assistance and Professor J. P. Scherft for critically reading the manuscript. This study was supported by the J. A. Cohen Institute, IRS, Leiden.

Received 3 February; accepted 6 March 1986.

1. Scheven, B. A. A., Burger, E. H., Kawilarang-de Haas, E. W. M., Wassenaar, A. M. & Nijweide, P. J. *Lab. Invest.* **53**, 72–79 (1985).
2. Marks, S. C. Jr *J. oral Path.* **12**, 226–256 (1983).
3. Visser, J. W. M., Bauman, J. G. J., Mulder, A. H., Eliason, J. F. & de Leeuw, A. M. *J. exp. Med.* **159**, 1576–1590 (1984).
4. Burger, E. H. *et al. J. exp. Med.* **156**, 1604–1614 (1982).
5. Thesingh, C. W. & Burger, E. H. *Dev. Biol.* **95**, 429–438 (1983).
6. Bol, S., Visser, J. & Van den Engh, G. *Expl. Hemat.* **7**, 541–553 (1979).
7. Bol, S. & Williams, N. *J. cell. Physiol.* **102**, 233–243 (1980).
8. Visser, J. W. M. & Bol, S. J. L. *Stem Cells* **1**, 240–249 (1981).
9. Baines, P., Bol, S. J. L. & Rosendaal, M. *Leukemia Res.* **6**, 81–88 (1982).
10. Till, J. E. & McCulloch, E. A. *Radiat. Res.* **14**, 213–222 (1962).
11. Lahiri, S. K., Keizer, H. J. & van Putten, L. M. *Cell Tissue Kinet.* **3**, 355–362 (1970).
12. Monette, F. C. & DeMello, J. B. *Cell Tissue Kinet.* **12**, 161–175 (1979).
13. Dorssers, L., Burger, E. H. & Wagemaker, G. *Expl. Hemat.* **12**, Suppl. 13, abstr. 3 (1984).
14. Kahn, A. J. & Simmons, D. J. *Nature* **258**, 325–327 (1975).
15. Walker, D. G. *J. exp. Med.* **142**, 651–663 (1975).
16. Ash, P., Loutit, J. F. & Townsend, K. M. S. *Nature* **283**, 669–670 (1980).
17. Loutit, J. F. & Nisbet, N. W. *Immunobiology* **161**, 193–203 (1982).
18. Visser, J. W. M., Bol, S. J. L. & Van den Engh, G. *J. Expl. Hemat.* **9**, 644–655 (1981).
19. Scheven, B. A. A., Kawilarang-de Haas, E. W. M., Wassenaar, A. M. & Nijweide, P. J. *Anat. Rec.* (in the press).

Epidermal growth factor-dependent phosphorylation of lipocortin

R. Blake Pepinsky & Lesley K. Sinclair

Biogen Research Corporation, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA

Lipocortin-like proteins are a family of steroid-induced inhibitors of phospholipase activity with potential anti-inflammatory activity (for reviews see refs 1, 2). Related proteins have been detected in a variety of tissues and species^{3–7}. The best characterized form is a protein of relative molecular mass (M_r) ~40,000 (40K), which is phosphorylated *in vivo* by protein tyrosine kinases and by protein

serine-threonine kinases^{8,9}. It has been proposed that the phospholipase inhibitory activity of lipocortin can be regulated by its phosphorylation⁷. In the A431 cell line, a protein of ~35K is phosphorylated by the protein tyrosine kinase activity of the epidermal growth factor (EGF) receptor¹⁰. Here we report that human lipocortin is phosphorylated near its amino terminus by the EGF receptor/kinase. By peptide mapping and immunological analyses, we show that lipocortin and the endogenous 35K substrate for the EGF receptor/kinase from A431 cells are the same protein.

Recently, the human gene for lipocortin was cloned and expressed in *Escherichia coli*¹¹. The gene encodes a 38.5K protein that inhibits phospholipase A₂ activity. The inhibitory activity was readily detected in crude bacterial lysates, but only from cells transformed with the human gene. Lipocortin was found in cell types of diverse origin, including macrophages, fibroblasts and epithelial cells⁴, and levels of expression in these preparations represented 0–0.5% of total cell protein. Figure 1 shows the results of gel filtration analysis of the recombinant human protein. On chromatography, the inhibitor yields a single peak with an apparent *M_r* of ~40K. The elution profile of phospholipase A₂ inhibitory activity (Fig. 1A) and the profile of the protein when subjected to SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 1B) coincide exactly. The inhibitory activity is trypsin-sensitive, destroyed by boiling and inactivated by treatment with heparin (not shown). Like natural lipocortin, the recombinant protein has a specific activity of 6,000–10,000 units per mg of protein, where 1 unit inhibits 15 ng of phospholipase A₂.

The EGF receptor (170K) is a protein tyrosine kinase that is activated by EGF^{12,13}. The enzyme is membrane-associated and retains its activity in preparations of isolated cell membranes^{14,15}. Two prominent phosphorylation products of the EGF-dependent kinase are the receptor itself and an endogenous 35K substrate. In the A431 cell line the receptor/kinase is overproduced¹⁶, leading to the frequent use of this human line as a source of the kinase. When membranes from A431 cells are incubated with [γ -³²P]ATP and the phosphorylated products analysed by SDS-PAGE, the receptor and the 35K protein are both phosphorylated^{14,15}. The degree of phosphorylation is greater if the kinase is prepared from EGF-treated cells. Using the A431 membrane-derived kinase activity, we observed similar results: both the receptor and the endogenous 35K protein were prominent phosphorylation products in the sample derived from EGF-treated membranes (Fig. 2b) and were present as minor species in the corresponding untreated sample (Fig. 2a).

When membranes from A431 cells were first treated with lipocortin and then subjected to phosphorylation and SDS-gel analysis, the recombinant protein was phosphorylated (Fig. 2c, d). Whereas in the absence of EGF, lipocortin was barely detected (Fig. 2c), it was the major phosphorylated protein in the EGF-treated preparations (Fig. 2d). The incorporation of phosphate into lipocortin varied as a function of the concentration of protein added (Fig. 2e–h). The dilution series tested ranged from concentrations where no effect was observed (5 nM lipocortin, Fig. 2e) to those where the phosphorylation reaction was saturated (5 μ M lipocortin, Fig. 2h). The increase in phosphorylation of lipocortin coincides with a decrease in phosphorylation of the EGF receptor. We found that human lipocortin and the endogenous 35K substrate have similar electrophoretic mobilities by SDS-gel electrophoresis. The 4K discrepancy between the predicted and apparent *M_r*s for lipocortin presumably results from its unusually charged amino-acid composition, as charged amino acids affect the electrophoretic mobility of proteins in SDS-gels¹⁷.

We have shown by peptide mapping analysis that lipocortin and the endogenous 35K protein from A431 cells are the same protein. We incubated gel slices containing phosphorylated lipocortin and phosphorylated 35K protein with CNBr, then separated the cleavage products by SDS-PAGE. Radioactive products were visualized by autoradiography. As shown in

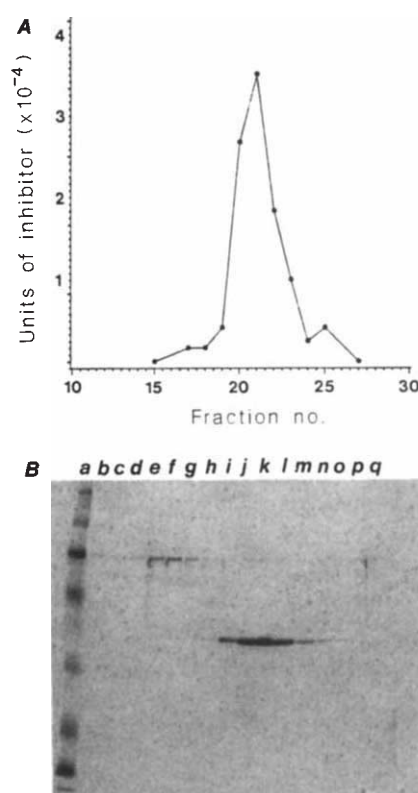


Fig. 1 Gel filtration analysis of recombinant lipocortin. Recombinant human lipocortin produced in *E. coli* was chromatographed on a P150 column (2.5 cm diameter $\times$ 45 cm). Aliquots from each 5-ml fraction collected were analysed for phospholipase A₂ inhibitory activity (A) and by SDS-PAGE (B). Proteins were separated electrophoretically in SDS-gels (15% acrylamide, 0.18% methylene bisacrylamide)²⁵ and stained with Coomassie brilliant blue R-250. Stacking gels contained 7.6% acrylamide and 0.21% bisacrylamide. a, Pre-stained *M_r* markers (BRL): myosin, 220K; phosphorylase B, 100K; bovine serum albumin, 65K; ovalbumin, 47K; α -chymotrypsinogen, 30K; β -lactoglobulin, 18K; lysozyme, 15K. Lanes b–q correspond to column fractions 10, 12, 14, 16–27 and 29, respectively. For inhibitory activity, samples were assayed with porcine pancreatic phospholipase A₂, using autoclaved ³H-oleic acid-labelled *E. coli* as substrate⁴.

Fig. 3b, c, phosphate-labelled CNBr fragments from the two protein preparations were identical. The six phosphorylated fragments formed a distinct subset of the total digestion products of lipocortin, which consists of 22 fragments (Fig. 3, lane a). This series of fragments is a mixture of complete and partial cleavage products, which we have characterized by CNBr mapping analysis. Although the relative intensities of the cleavage products vary as a function of the CNBr concentration, no additional products were generated when preparations were treated with CNBr at 7–21 mg ml⁻¹, spanning the range of concentrations used routinely for analysis of lipocortin.

To localize the site of phosphorylation within lipocortin, we first identified the fragment composition for each of its 22 digestion products by CNBr mapping and then used this information to identify the subset of fragments that were phosphorylated. This type of procedure has been described in detail previously^{18,19} and is based on the fact that each CNBr fragment is represented by a distinct subset of the digestion products. The mapping information for lipocortin is summarized on the left of Fig. 3; however, the actual cleavage maps will be presented elsewhere. Based on the map of lipocortin, the pattern of phosphorylated fragments suggests a modification within fragment 1. The series of radioactive spots corresponds to fragment 1 and each of the fragment 1-containing partial digests. Thus, the

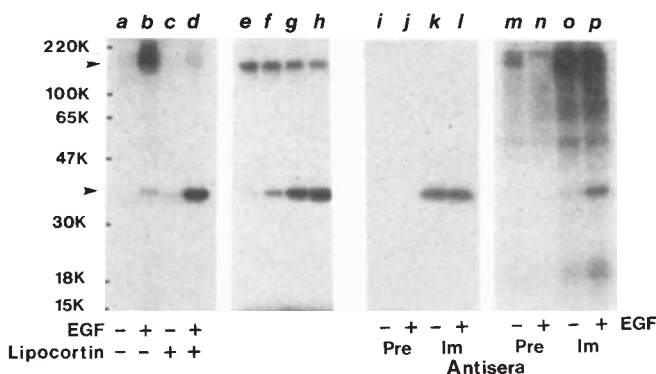


Fig. 2 EGF-dependent phosphorylation of lipocortin. A431 cells were grown at 37 °C in Dulbecco's modified Eagle's medium supplemented with 10% calf serum. *In vitro* phosphorylation reactions using isolated A431 cell membranes (lanes a–h) and immune precipitations from metabolically labelled cells (^{35}S -methionine-labelled, lanes i–l; ^{32}P -phosphate-labelled, lanes m–p) were analysed by SDS-PAGE and radioactive products visualized by autoradiography. Arrowheads denote the positions of the EGF receptor and the 35K protein. e–h, *In vitro* phosphorylation products generated when membranes from EGF-treated cells were treated with various concentrations of lipocortin: 0.2, 2, 20 and 200 $\mu\text{g ml}^{-1}$ of lipocortin, respectively.

Methods. For *in vitro* phosphorylation reactions, confluent monolayers of A431 cells in T75 flasks were washed with Dulbecco's medium without serum and incubated at 37 °C for 1 h in 7 ml of serum-free medium without (–) and with (+) 200 ng ml^{-1} EGF (Sigma). Monolayers were rinsed with 2.5 ml of ice-cold hypotonic buffer (10 mM HEPES pH 7.5, 0.5 mM MgCl_2) and the cells scraped into 1.5 ml of the same buffer. After 30 min on ice, the cells were disrupted with a Dounce homogenizer (20 strokes with a type A pestle), and large particulate debris was removed by centrifugation (2,000g, 5 min). The resultant supernatants were centrifuged at 10,000g for 10 min and the membranous pellets suspended in 300 μl of phosphorylation buffer (20 mM HEPES pH 7.5, 2 mM MgCl_2 , 10 μM ammonium vanadate). EGF-treated membranes were suspended in buffer plus 2 $\mu\text{g ml}^{-1}$ EGF. 60- μl reactions, each containing 30 μl of a given membrane preparation, were held on ice for 10 min in the presence of 30 μCi [γ - ^{32}P]ATP per tube. Reactions were stopped by adding 20 μl of 5 $\times$ electrophoresis sample buffer. Samples were heated at 55 °C for 10 min, and then subjected to SDS-gel analysis. For the metabolic labelling experiments, A431 cells were grown in 6-well microtitre dishes. Cells were labelled with ^{35}S -methionine (100 μCi per well) in methionine-free medium and with ^{32}P -phosphate (1 mCi per well) in phosphate-free medium. In both cases cells were cultured for 4 h in serum-free medium without and with 200 ng ml^{-1} EGF. Labelled cells were disrupted with 1 ml of lysis buffer (50 mM NaCl, 20 mM Tris-HCl pH 7.3, 0.5% Nonidet P-40, 0.5% sodium deoxycholate, 4 mM iodoacetic acid, 1 mM ammonium vanadate), and lysates were subjected to vortex mixing for 30 s and clarified by centrifugation for 5 min in an Eppendorf centrifuge. The appropriate antisera were added (Pre, preimmune serum; Im, immune serum) and the samples were incubated for 2 h at 4 °C. The immune complexes were collected by adsorption to protein A-Sepharose for 1 h at 4 °C with continuous mixing. The adsorbed immune precipitates were washed four times (each with 0.5 ml lysis buffer), suspended in 50 μl of electrophoresis sample buffer, heated at 60 °C for 10 min, and subjected to SDS-PAGE. To prepare anti-lipocortin serum, rabbits were immunized with recombinant human protein produced in *E. coli*. Native protein was emulsified with Freund's complete adjuvant and administered by intramuscular injection; 5 weeks after the primary injection, the rabbits received a booster injection of protein emulsified with incomplete adjuvant.

phosphorylation site is near the amino terminus of lipocortin. The positions of serine, threonine and tyrosine residues within fragment 1 place further boundaries on this site. Recombinant lipocortin also can be phosphorylated *in vitro* by the catalytic subunit of protein kinase A. The cleavage profile of protein

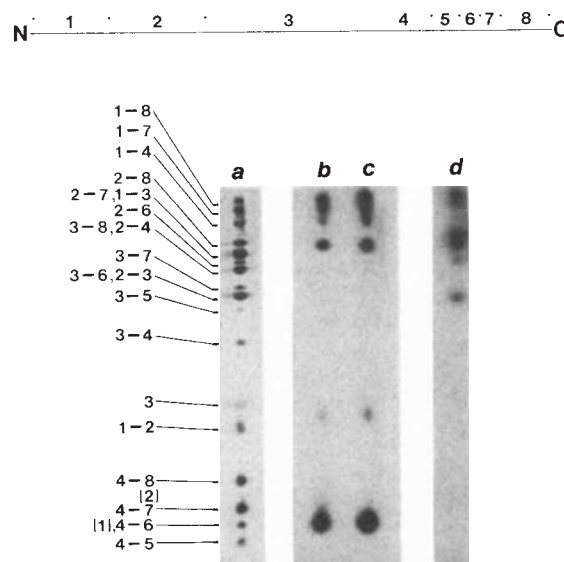


Fig. 3 Cyanogen bromide mapping of phosphorylated lipocortin. a, Unlabelled lipocortin; b, phosphorylated lipocortin (EGF receptor/kinase); c, phosphorylated endogenous 35K protein (EGF receptor/kinase); d, phosphorylated lipocortin (protein kinase A). Numbers on the left correspond to CNBr fragments of lipocortin: fragment 1 refers to the amino-terminal fragment and fragment 8 the carboxy-terminal fragment. The distribution and relative sizes of all the fragments are indicated in the schema shown at the top. The fragmentary compositions of cleavage fragments are based on CNBr mapping¹⁸. Although fragments 1 and 2 are not visualized by this particular silver staining procedure²⁶, their positions in cleavage profile are indicated. Other cleavage products were not identified because either they were below the size limit of detection of this particular gel system or they were minor cleavage products and could not be analysed.

Methods. Gel slices containing either unlabelled lipocortin or *in vitro* phosphorylated products were incubated with CNBr (10 mg ml^{-1}) for 1 h at room temperature in 0.1 M HCl plus 0.1% 2-mercaptoethanol. The CNBr-treated gel slices were washed twice for 5 min each with water, once for 5 min in 0.25 M Tris-HCl pH 6.8, and once for 10 min at 37 °C in electrophoresis sample buffer. The slices were loaded onto SDS-gels and the cleavage products separated by SDS-PAGE using the Laemmli discontinuous buffer system (main gel: 15% acrylamide, 0.4% methylene bisacrylamide).

kinase A-phosphorylated lipocortin (Fig. 3d) suggests a modification within fragment 3.

In addition to the *in vitro* experiments, we have evaluated the phosphorylation state of lipocortin in metabolically labelled A431 cells. Figure 2i–p shows the results of an analysis in which crude lysates from labelled cells were treated with either anti-lipocortin serum or preimmune serum, and the immune complexes analysed by SDS-PAGE. In lysates of ^{35}S -methionine-labelled cells, the anti-lipocortin serum recognized a single protein of 35K which was not detected with preimmune serum from the same rabbit. Metabolic labelling of the 35K protein was independent of EGF treatment. The 35K protein was also detected in immune precipitates from ^{32}P -phosphate-labelled cell lysates (Fig. 2m–p), but only for preparations derived from EGF-treated cells. Similar results for metabolic labelling of the endogenous 35K protein have been described previously¹⁰. Using conditions that enrich for tyrosine phosphate (that is, the gel is treated prior to autoradiography with 1 M KOH for 1 h at 55 °C; ref. 20) we observed that the phosphorylated 35K protein was enhanced relative to background radioactivity (not shown).

Here we have presented data demonstrating that the endogenous 35K substrate for the EGF receptor/kinase from A431 cells is lipocortin, and therefore presumably a potent phospholipase A_2 inhibitor. This result was deduced from both

structural and immunological analyses. The relevance of the phospholipase inhibitory activity to the cellular function of these proteins remains to be established. Peptide mapping and the use of the anti-lipocortin antiserum should allow us to clarify rapidly the relationships between the 35K substrate for the EGF-dependent kinase^{15,21,22}, the 34–40K substrate for *src*-like kinases^{21,23,24} and the 35K G-protein β -subunit²², which to date remain controversial. The role of the 35K protein as a phospholipase A₂ inhibitor probably represents an important biological activity in EGF transduction.

We thank Barbara P. Wallner for the JA221 strain of *E. coli* transformed with plasmid LipTrc155A, a strain producing recombinant human lipocortin; and John Smart, Vicky Sato, Richard Flavell and Jeff Browning for critical reading of the manuscript. This work was funded by Biogen Research Corporation and by Yamanouchi Pharmaceutical Company.

Received 6 February; accepted 20 March 1986.

1. Flower, R. J., Wood, J. N. & Parente, L. *Adv. Inflamm. Res.* **7**, 61–69 (1984).
2. Hirata, F. *Adv. Inflamm. Res.* **7**, 71–78 (1984).
3. Blackwell, G. J. *et al. Br. J. Pharmac.* **76**, 185–194 (1982).
4. Pepinsky, R. B. *et al. J. Biol. Chem.* **261**, 4239–4246 (1986).
5. Hirata, F. *J. Biol. Chem.* **256**, 7730–7733 (1981).
6. Rothhut, B., Russo-Marie, F., Wood, J., DiRosa, M. & Flower, R. J. *Biochem. biophys. Res. Commun.* **117**, 878–884 (1983).
7. Gupta, C., Katsumata, M., Goldman, A. S., Herold, R. & Piddington, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1140–1143 (1984).
8. Hirata, F. *et al. Biochem. biophys. Res. Commun.* **118**, 682–690 (1984).
9. Hirata, F., Matsuda, K., Notsu, Y., Hattori, T. & DelCarmine, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4717–4721 (1984).
10. Sawyer, S. T. & Cohen, S. *J. Biol. Chem.* **260**, 8233–8236 (1985).
11. Wallner, B. P. *et al. Nature* **320**, 77–81 (1986).
12. Ushiro, H. & Cohen, S. *J. Biol. Chem.* **255**, 8363–8365 (1980).
13. Cohen, S., Ushiro, H., Stoscheck, C. & Chinkers, M. *J. Biol. Chem.* **257**, 1523–1531 (1982).
14. King, L. Jr, Carpenter, G. & Cohen, S. *Biochemistry* **19**, 1524–1528 (1980).
15. Fava, R. A. & Cohen, S. *J. Biol. Chem.* **259**, 2636–2645 (1984).
16. Fabricant, R. N., DeLarco, J. E. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 565–569 (1977).
17. Tung, J. S. & Knight, C. A. *Analyt. Biochem.* **48**, 153–163 (1972).
18. Pepinsky, R. B. *J. Biol. Chem.* **258**, 11229–11235 (1983).
19. Pepinsky, R. B. & Vogt, V. M. *J. Virol.* **52**, 145–153 (1984).
20. Cooper, J. A., Sefton, B. M. & Hunter, T. *Meth. Enzym.* **99**, 387–402 (1983).
21. Glenney, J. R. & Tack, B. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7884–7888 (1985).
22. Valentine-Braun, K. A., Northup, J. K. & Hollenberg, M. D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 236–240 (1986).
23. Gerke, V. & Weber, K. *J. Biol. Chem.* **260**, 1688–1695 (1985).
24. Erikson, E. & Erikson, R. L. *Cell* **21**, 829–836 (1980).
25. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
26. Wray, W., Boulakas, T., Wray, V. P. & Hancock, R. *Analyt. Biochem.* **118**, 197–203 (1981).

Ha-ras hypervariable alleles in myelodysplasia

S. L. Thein*, D. G. Oscier†, J. Flint* & J. S. Wainscoat‡

*MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, and ‡Department of Haematology, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK
†Department of Haematology, Royal Victorian Hospital, Boscombe, Bournemouth BH1 4JG, UK

The somatic mutation of one of the *ras* oncogenes is now considered to be a critical step in the pathogenesis of many tumours^{1,2}. Circumstantial evidence also suggests that some individuals may be genetically predisposed to malignancy³ and a general method used to analyse such disease susceptibility is the study of restriction fragment length polymorphisms (RFLPs) at particular loci^{4,5}. The Harvey *ras* (Ha-*ras*) locus includes a hypervariable region (HVR) which consists of a series of 28-base-pair (bp) tandem repeats 3' to the gene⁶. This arrangement gives rise to alleles of a wide range of sizes, making such genetic analysis possible. A previous study reported that white blood cell DNA from cancer patients frequently showed allelic restriction fragments at the Ha-*ras* locus which were found only rarely in normal unaffected individuals, and it was concluded that the inheritance of such unusual alleles may

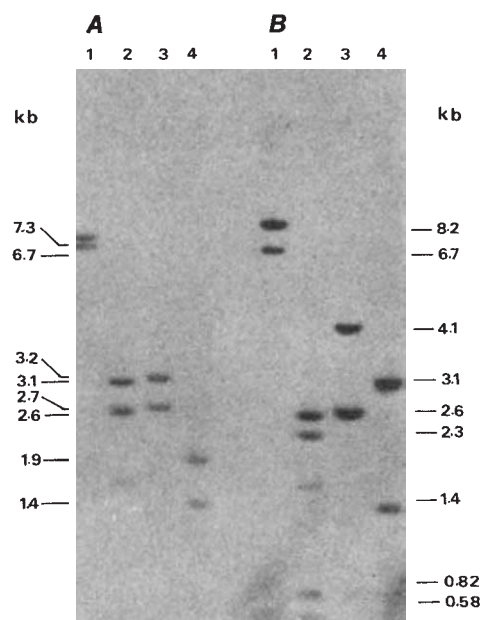


Fig. 1 Comparison of the Ha-*ras* alleles generated by four restriction endonucleases. DNAs from two individuals (A and B) were digested with *Bam*HI (lane 1), *Taq*I (lane 2), *Hinf*I (lane 3) and *Ava*II (lane 4). The digested DNA was hybridized to the 2.5-kb *Bam*HI-*Sst*I fragment containing the T24/EJ bladder oncogene⁶. A, An individual with the *a/b* Ha-*ras* alleles corresponding to the *Bam*HI 6.7/7.3-kb fragment and the *Taq*I 2.6/3.1-kb, *Hinf*I 2.7/3.2-kb and *Ava*II 1.4/1.9-kb fragments. B, An individual with two *Taq*I site polymorphisms. Corresponding sizes for the *a* allele were seen with *Bam*HI (6.7 kb), *Taq*I (2.6 kb), *Hinf*I (2.7 kb) and *Ava*II (1.4 kb) digests. *Bam*HI, *Hinf*I and *Ava*II digests gave appropriate corresponding sizes (8.2, 4.1 and 3.1 kb respectively) for the 'd' allele but a pattern of three fragments (2.3 kb, 820 bp and 580 bp) was obtained for the *Taq*I digests. Further investigations suggested the presence of two *Taq*I site polymorphisms within the HVR. All alleles with the *Taq*I site polymorphisms were characterized by the *Ava*II 'd' (3.1 kb) alleles, which suggests that these *Taq*I site polymorphisms are linked to a particular HVR allele. **Methods.** DNA was prepared from peripheral blood buffy coat cells by digestion with proteinase-K and phenol-chloroform extraction as described previously¹⁰. DNA (5 µg) was digested with 25 units of each restriction enzyme, separated by electrophoresis on a 0.8% agarose gel and transferred to nitrocellulose membrane by Southern blotting. The *Bam*HI-*Sst*I fragment of plasmid pEJ⁶ containing the Ha-*ras* HVR was purified and nick-translated to a specific activity of $\sim 1.0 \times 10^8$ c.p.m. µg⁻¹. Pre-hybridization, hybridization and washing of the membranes were done as described previously¹⁰. Autoradiography was at -70 °C.

be linked to a susceptibility to cancer⁷. As this conclusion has major implications we sought to investigate whether this association could be confirmed in patients with myelodysplasia, a common haematological malignancy reported to have the highest frequency of rare alleles⁷. The Ha-*ras* alleles were characterized in normal healthy individuals and compared with those found in patients with myelodysplasia (MDS). Our results, reported here, show that the distribution of Ha-*ras* alleles in myelodysplastic patients is not significantly different from that in normal individuals.

Fifty-two normal healthy individuals and 53 patients with MDS were studied. The diagnosis of MDS fulfilled the criteria of the FAB classification⁸. DNA was extracted from peripheral blood leukocytes, digested with appropriate restriction enzymes, Southern-blotted and hybridized to the *Bam*HI-*Sst*I fragment of plasmid EJ containing the HVR~1.5 kilobases (kb) from the 3' terminus of the coding sequences of the Ha-*ras* gene. The first objective was to choose the most appropriate restriction enzyme to give maximal resolution of the different HVR alleles.

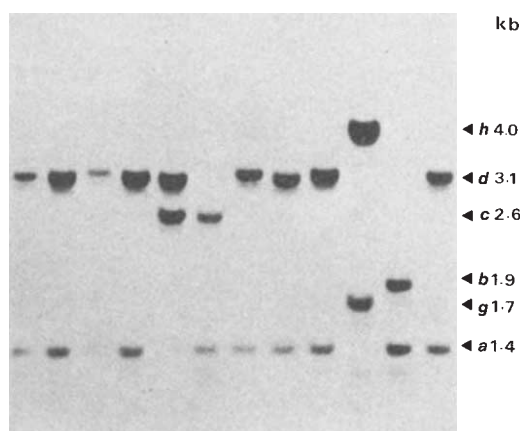


Fig. 2 Microheterogeneity of the common Ha-ras allele. DNA (5 µg) was digested with *Ava*II, electrophoresed on a 0.8% agarose gel, and transferred to nitrocellulose membrane by Southern blotting. Hybridization and autoradiography were as described in Fig. 1 legend. Microheterogeneity between the 'd' alleles could only be visualized when the DNA samples were run in adjacent tracks. However, the differences between the various 'd' alleles were too small to be measured reliably. There was no obvious microheterogeneity among the different 'a' alleles.

Table 1 Distribution of the *Ava*II Ha-ras alleles

Allele	Size (kb)	Normal individuals	Myelodysplastic individuals	Total
Common				
a	1.4	71 (68%)	66 (62%)	137 (65%)
b	1.9	6 (6%)	15 (14%)	21 (10%)
c	2.6	9 (9%)	14 (13%)	23 (11%)
d	3.1	13 (13%)	8 (8%)	21 (10%)
Rare				
e	2.0	0 (0%)	0 (0%)	0 (0%)
f	1.5	3 (3%)	3 (3%)	6 (3%)
g	1.7	1 (1%)	0 (0%)	1 (<1%)
h	4.0	1 (1%)	0 (0%)	1 (<1%)
Total		104	106	210

DNA was prepared from peripheral blood buffy coat cells and the allele sizes determined as described in Fig. 1 legend. The numbers of particular *Ava*II Ha-ras alleles are given with the frequencies in parentheses.

DNAs from 10 unrelated individuals were digested with a panel of restriction enzymes and resolution of the different bands was found to be greatest with the enzyme *Ava*II, which cleaves just outside the HVR. In addition to *Ava*II, all DNAs were digested with *Bam*HI and *Taq*I to ensure that variation in allelic fragment sizes was directly related to the HVR length rather than to restriction enzyme site polymorphisms (see Fig. 1).

Although optimal resolution was obtained with *Ava*II digestion, it was difficult to compare similar-sized Ha-ras alleles between different gels. To overcome this problem, all DNAs with apparently similar-sized alleles were re-electrophoresed in adjacent tracks against control DNA samples (Fig. 2). Nine measurably different *Ava*II Ha-ras alleles were found, of which four were common and five rare. Microheterogeneity of fragment sizes was observed within the four common allele classes in both groups of individuals. However, the degree of microheterogeneity was so small that it could be visualized only by examination of the autoradiographs when alleles of the same size class were in adjacent tracks; these small differences cannot be measured reliably. The degree of microheterogeneity was greater within the three classes of common, larger-sized alleles

than within the smallest common allele class. This phenomenon probably relates to the greater likelihood of misalignment occurring during recombination between larger HVR alleles. This result is in accord with the observation in a study of the hypervariable minisatellite regions that those probes with the longer HVRs detected more allele sizes⁹. There was no observable difference in the degree of microheterogeneity in the allele sizes between the two groups of individuals.

Table 1 shows the numbers of the different Ha-ras alleles found in the two groups of individuals. The frequencies of the four common *Ava*II alleles are not significantly different between the two groups and are very similar to those previously reported using *Bam*HI. The rare alleles were found sporadically in both groups: 5/104 in normal chromosomes and 3/106 in MDS chromosomes. Therefore, there is no significant increase in the frequency of rare alleles in the MDS patients compared with normal individuals. The frequency of the rare alleles in our group of normal individuals (4.8%) was not significantly different from that in the group of normal individuals observed previously (3.9%)⁷. The patients reported previously as having the highest proportion of rare alleles had MDS, 6/7 patients having rare alleles⁷. Here we studied 53 patients with MDS of which only three had rare alleles. Therefore, our finding that the MDS patients do not have an increased frequency of rare alleles is in striking contrast to the previous study, although the reasons for this difference are not clear. Thus, we have found no evidence to suggest that the inheritance of rare HVR alleles at the Ha-ras locus predisposes to myelodysplasia.

We thank Liz Gunson for typing the manuscript; Drs D. R. Higgs and A. J. Jeffreys for critical comments on the manuscript; and Dr F. G. Bolton and Professor D. J. Weatherall for their support and encouragement.

Received 3 February; accepted 7 March 1986.

1. Bos, J. L. *et al.* *Nature* **315**, 726-730 (1985).
2. Bos, J. L. *et al.* *Nucleic Acids Res.* **12**, 9155-9163 (1984).
3. Burt, R. W. *et al.* *New Engl. J. Med.* **312**, 1540-1544 (1985).
4. Gusella, J. F. *et al.* *Nature* **306**, 234-239 (1983).
5. Thein, S. L. & Wainscoat, J. S. *Disease Markers* (Wiley, New York, in the press).
6. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. *Nature* **302**, 33-37 (1983).
7. Krontiris, T. G., Di Martino, N. A., Colb, M. & Parkinson, D. R. *Nature* **313**, 369-373 (1985).
8. Bennett, J. M. *et al.* *Br. J. Haemat.* **51**, 189-199 (1982).
9. Jeffreys, A. J., Wilson, J. & Thein, S. L. *Nature* **316**, 67-73 (1985).
10. Old, J. M. & Higgs, D. R. in *Methods in Haematology*, vol. 6, 74-102 (Churchill Livingstone, Edinburgh, 1983).

Spectrin assembly in avian erythroid development is determined by competing reactions of subunit homo- and hetero-oligomerization

Catherine M. Woods & Elias Lazarides

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

Erythroid differentiation entails the biogenesis of a membrane skeleton, a network of proteins underlying and interacting with the plasma membrane¹⁻³, whose major constituent is the heterodimeric protein spectrin, composed of two structurally similar but distinct subunits⁴, α (relative molecular mass (M_r) 240,000) and β (M_r 220,000), which interact side-on with each other to form a long rod-like molecule^{1,3,5-10}. Interaction of this network with the membrane is mediated by the binding of the β subunit to ankyrin^{1,6,11-14}, which in turn binds to the cytoplasmic domain of the transmembrane anion transporter (also referred to as band 3)^{1,15-19}. Purified α and β subunits of spectrin from the membrane of mature red blood cells will spontaneously

heterodimerize^{13,14,20-23}, suggesting that assembly of the spectrin-actin skeleton is a simple self-assembly process, but *in vivo* studies with developing chicken erythroid cells have indicated that assembly *in vivo* is more complex²⁴⁻²⁹. We now present evidence that newly synthesized spectrin subunits *in vivo* or *in vitro* rapidly adopt one of two competing conformations, a heterodimer or a homo-oligomer. These competing reactions seem to determine the overall extent of spectrin assembled during erythroid development by determining which conformation will assemble onto the membrane-skeleton (the heterodimer) and which conformations are targeted for degradation (the homo-oligomers).

When chicken erythroid cells are pulse labelled with ³⁵S-methionine and lysed hypotonically, excess unassembled ankyrin, β -spectrin and α -spectrin remain soluble²⁹. We characterized the unassembled polypeptides by chromatographing hypotonically soluble fractions through an Ultrogel Ac22 gel filtration column, and found that unassembled β - and α -spectrin eluted as distinct species (Fig. 1A), well ahead of the fractions where globular proteins of their size would be expected to appear. β -Spectrin was only just included in the column, eluting with the purified adult red blood cell $\alpha\beta$ -spectrin heterotetramers used as a standard eluate. Most of the α -spectrin migrated behind β -spectrin, at the position of the heterodimeric red blood cell spectrin standard. Ankyrin eluted in a manner that would be predicted for a globular monomeric protein of its size, slightly ahead of haemoglobin and the bulk of the other polypeptides.

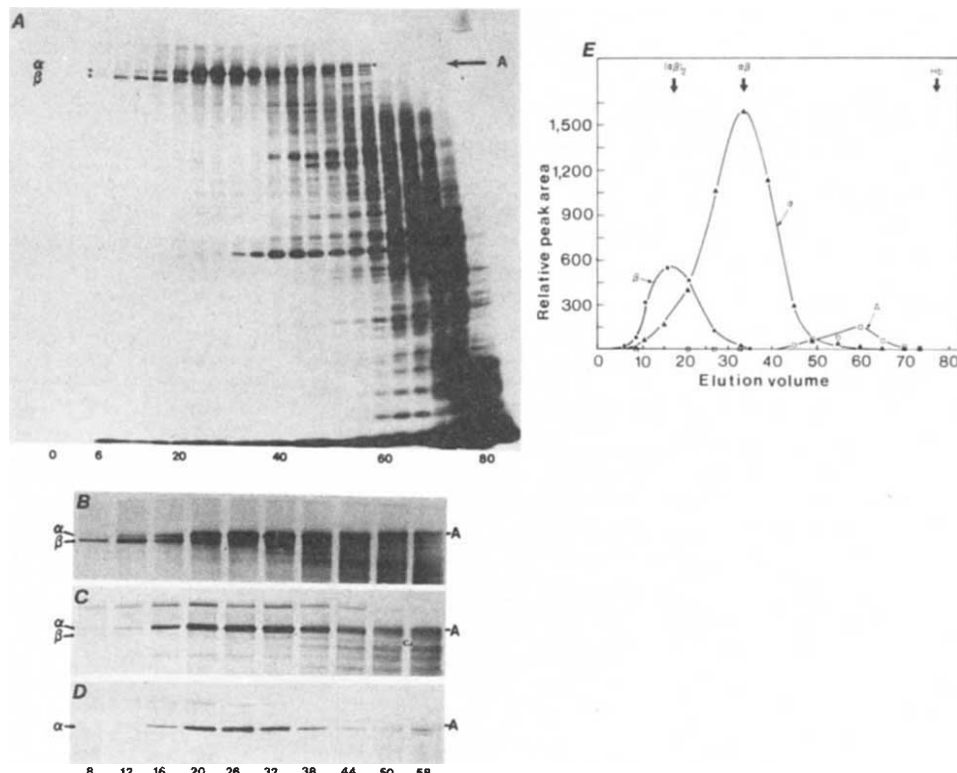
Immunoprecipitation of unassembled β -spectrin, α -spectrin

and ankyrin across the column after a pulse-labelling period as short as 5 min showed clearly that the three polypeptides eluted as distinct peaks, with most β - and α -spectrin eluting respectively with the $\alpha\beta$ -heterotetramer and $\alpha\beta$ -heterodimer standards (Fig. 1E). The high mobility of the unassembled β -spectrin was not the result of association with organelles because these should have eluted in the void volume and this β -spectrin species was clearly included in the column. Also, pretreatment of the hypotonically soluble fraction with puromycin, to release nascent polypeptides from ribosomes, or inclusion of 0.5% Triton X-100 in the elution buffer, to disrupt membranous organelles, had no effect on the elution pattern of either α - or β -spectrin (data not shown). Furthermore, neither unassembled α - or β -spectrin sedimented after centrifugation at 150,000g for 30 min (see Table 1). Thus, β -spectrin seems to be associated in a species of relative molecular mass equivalent to that of a spectrin tetramer and α -spectrin to that of a spectrin dimer.

Further evidence that the unassembled β - and α -spectrin subunits were indeed separate homo-oligomeric species was obtained by analysing fractions off the column by two-dimensional non-denaturing followed by SDS-polyacrylamide gel electrophoresis (PAGE). The two spectrin species clearly did not co-migrate even in fractions where the β - and α -spectrin peaks overlapped; β -spectrin consistently migrated close to the position of adult red blood cell spectrin $\alpha\beta$ -heterotetramers and α -spectrin with spectrin $\alpha\beta$ -heterodimers (Fig. 2A,B). Furthermore, no other protein co-migrating with either spectrin homo-oligomeric species was observed (Fig. 2A) even after labelling

Fig. 1 Gel-filtration elution profiles of the hypotonically soluble fraction of pulse-labelled erythroid cells from 10-day chick embryos. Results shown are for lysates from erythroid cells lysed immediately after 10 min labelling with ³⁵S-methionine (A), after 10 min pulse-labelling followed by 30 min chase (B), after 1 h chase (C), or 4 h chase in the presence of cold methionine (D) and quantitation of β -spectrin, α -spectrin and ankyrin in fractions across the column by immunoprecipitation (E). A-D show autoradiograms after exposure for 48 h. α = α -spectrin; β = β -spectrin; A = ankyrin. The numbers beneath the lanes represent V_e , that is, volumes eluting after the void volume. E, Quantitation by immunoprecipitation of β -spectrin (β), α -spectrin (α) and ankyrin (A) across the column shown in A. B The arrows indicate the peak fractions of the rabbit spectrin ($\alpha\beta$)₂-tetramer and $\alpha\beta$ -dimer, respectively, and also the fraction where haemoglobin begins to appear (Hb).

Methods. Samples (0.75 ml) of packed, washed erythroid cells isolated from 10-day chick embryos as described previously²⁹ were preincubated in methionine-free minimal essential medium (MEM) supplemented with 5% dialysed fetal calf serum (FCS) labelled for 10 min with 300 μ Ci ³⁵S-methionine (specific activity 11,000-13,000 Ci mmol⁻¹) and chased for various time intervals in fresh MEM-FCS media containing 2 mM unlabelled methionine all at 38 °C in a humidified atmosphere. Aliquots of cells were collected by centrifugation in ice-cold choline chloride (155 mM), HEPES (5 mM) buffer pH 7.2, then lysed in 4 vols ice-cold hypotonic lysis buffer (5 mM Tris-HCl pH 7.2, 5 mM MgCl₂, 0.1 mM dithiothreitol (DTT), 1 mM EGTA, 1 mM *p*-tosyl-L-arginine methyl ester (TAME) 1 mM phenylmethylsulphonyl fluoride). After centrifugation of the cell lysate at 13,000g for 10 min in an Eppendorf centrifuge, the supernatant was recentrifuged for 10 min before being layered on a 2.5 × 50 cm Ac22 Ultrogel gel filtration column (exclusion limit 3 × 10⁶ daltons) and eluted with Tris-buffered saline (125 mM NaCl, 10 mM Tris-HCl pH 7.2, 1 mM MgCl₂, 1 mM EDTA, 1 mM NaH₂PO₄, 0.1 mM DTT; identical elution patterns of the spectrins and ankyrin were obtained whether or not Mg²⁺ or DTT was included) at a flow rate of 15 ml h⁻¹. Beginning at the void volume, the eluate was collected in 2-ml fractions and alternate fractions were assayed by SDS-PAGE using the modified Laemmli gel system described by Hubbard and Lazarides³⁶. After processing in 20% diphenyloxazole (PPO) in dimethyl sulphoxide, dried gels were autoradiographed with Kodak XAR film at -80 °C. The levels of β -spectrin, α -spectrin and ankyrin were quantitated by immunoprecipitation using method B of Blikstad *et al.*²³, except that deoxycholate was omitted from the buffer for ankyrin immunoprecipitation, followed by SDS-PAGE. The gels were processed for autoradiography, dried and exposed to pre-flashed Kodak XAR film long enough to give bands of intensity within the linear range of intensity versus amount of ³⁵S. The bands were then scanned with a Quick Scan (Helena Laboratories) and the peaks integrated by computer. Values represent arbitrary relative units. Rabbit spectrin $\alpha\beta$ -heterotetramers and heterodimers were isolated from rabbit erythrocyte ghosts by the method of Ungewickell and Gratzer²⁰. Their elution profiles through the column were assayed by SDS-PAGE followed by densitometry of the Coomassie-stained gels.



periods of 3–4 h (data not shown). Crosslinking experiments with dithiobis(succinimidyl)propionate provided further evidence that unassembled α -spectrin existed as a homodimer, for it could be efficiently converted to a species of $M_r \approx 0.5 \times 10^6$, while β -spectrin existed as a homotetramer since an additional product of $M_r \approx 10^6$ was observed on one-dimensional SDS gels in non-reducing conditions (Fig. 2C).

These elution patterns of the unassembled pulse-labelled spectrins and ankyrin persisted until the polypeptides had been turned over (Fig. 1B–D), the rates of disappearance of the β -spectrin, α -spectrin and ankyrin peaks matching the individual rates of turnover previously described for these polypeptides ($t_{1/2} = 15$ –20 min, 2 h and 1 h, respectively²⁹). Together, our observations indicate that, unlike the assembled polypeptides, unassembled β -spectrin, α -spectrin and ankyrin behave as independent kinetic species and exhibit no obvious association with each other.

Isolated α - and β -spectrin subunits purified from adult red blood cell heterodimeric spectrin after denaturation with urea or SDS reassociate spontaneously into heterodimers upon renaturation^{13,14,20,30}. Any α - and β -homo-oligomeric species that might form as the separated subunits are renatured have been considered to be metastable species that will spontaneously convert into heterodimers. The unassembled β -homotetramer and the α -homodimer behaved quite differently in this respect, as there was no evidence of interconversion to an $\alpha\beta$ -heterodimer or tetramer even after incubation of an approximately stoichiometric mixture of the β - and α -spectrin peak fractions for 1 h at 37 °C and analysis by sieving chromatography. However, if such a mixture was first denatured with 3 M urea or 0.5% SDS and allowed to renature before chromatography²⁰, α - and β -spectrin co-eluted. Analysis by two-dimensional-PAGE confirmed that hetero-oligomerization occurred only after the two spectrin homo-oligomers had been previously denatured (data not shown). These results suggest that the homo-oligomeric forms assumed by the newly synthesized α - and β -spectrin subunits represent alternative, but stable, conformations to the normal ($\alpha\beta$ -heterodimer)_n form of assembled spectrin. Similar association experiments showed no evidence of association between the unassembled homotetrameric β -spectrin and unassembled ankyrin (data not shown).

We next tested whether the unassembled forms of spectrin could bind to the erythroid plasma membrane, and whether binding would alter their capacity to interact with each other, by using spectrin- or ankyrin-depleted erythrocyte inside-out membrane vesicles (IOVs), which have been shown to bind chicken spectrin in a similar manner to rabbit spectrin³¹. The β -spectrin homo-tetramer was found to bind selectively to spectrin-depleted IOVs and this binding was inhibited by prior removal of ankyrin or prior trypsinization of the vesicles (Table 1a). In contrast, the α -spectrin homodimer did not bind to IOVs, whether added alone or in the presence of stoichiometric amounts of the β -spectrin homotetramer (Table 1a). The possibility that filtration through the Ultrogel column had somehow affected the membrane-binding properties of these unassembled spectrin forms was discounted because similar results were obtained with membrane-binding studies on the hypotonically soluble fraction immediately after cell lysis (Table 1b). Some binding of the unassembled ankyrin was also detected but this was often accompanied by proteolysis. The inability of the α -spectrin homodimer to bind to spectrin-depleted IOVs, even in the presence of the β -spectrin homotetramer, eliminates the possibility that heterodimerization occurs by nucleation of homo-oligomers by the membrane.

Previous kinetic analysis has indicated that the unassembled pools of the two spectrin subunits represent ~50% of the β -spectrin and 85% of the α -spectrin synthesized after a 10-min pulse-labelling period^{24,25}. The rest assembles rapidly with the stoichiometry (1:1) observed at steady state (refs 24, 25 and

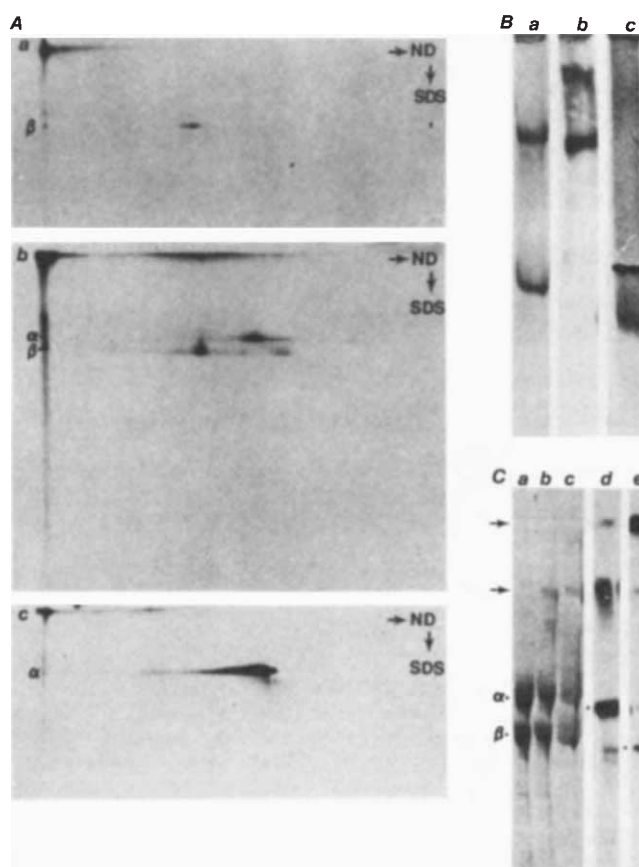


Fig. 2 **A, B**, Analysis of the β -spectrin and α -spectrin species from the β -peak fraction (**Aa, Bb**), the intermediate fraction (**Ab**) and the α -peak fraction (**Ac, Bc**) by non-denaturing PAGE in the first dimension (**B**) followed by SDS-PAGE in the second dimension (**A**), showing that the soluble unassembled α -spectrin and β -spectrin are distinct and separate species. Aliquots identical to those fractions eluting at 8 ml, 18 ml and 28 ml in Fig. 1A were analysed by PAGE on a 2.5–5% gradient, polyacrylamide gel under non-denaturing conditions (ND) as described by Morrow and Marchesi²² in the first dimension, with duplicate lanes being excised and run in a second dimension on 10% SDS-PAGE (**Aa, b, c**, respectively). Gels were stained with Coomassie blue and then destained before being processed for autoradiography (PPO for the SDS gels, En³HANCE for the non-denaturing gels), dried and exposed to Kodak XAR at –80 °C for 5 days (non-denaturing gels) or 18 days (second-dimension SDS gels). Lane **a** in panel **B** is the Coomassie-stained lane of rabbit spectrin heterotetramers and heterodimers run as standards on the non-denaturing gel. Lane **b** represents the β -spectrin peak fraction ($V_e = 8$ ml) and lane **c** the α -spectrin fraction ($V_e = 28$ ml). The smear running diagonally down to the right-hand side from the α -spectrin spot represents α -spectrin breakdown products and/or nascent chains, since they react with α -spectrin-specific antibodies under stringent conditions but not with β -spectrin antibodies. Some aggregation occasionally occurs in both rabbit spectrin standards and the radioactive chicken spectrin species on electrophoresis in the first (native) dimension and is seen as material at the top of the gel. **C**, The unassembled α -spectrin and β -spectrin species can be crosslinked to a species of $M_r \approx 0.5 \times 10^6$ and $\sim 1 \times 10^6$, respectively. Rabbit spectrin heterodimer standards (**a–c**) and aliquots from the peak fractions of hypotonically soluble β -spectrin (**d**) and α -spectrin (**e**), partially purified by gel filtration, were incubated in 10^{-3} M dithiobis(succinimidyl)propionate for 0 min (**a**), 10 min (**b**) or 30 min (**c–e**) at 25 °C. The samples were then denatured by addition of 1/5th volume 5% SDS sample buffer in the absence of any reducing agent and boiled for 5 min before loading onto a 3.5–10% polyacrylamide gradient SDS gel. After electrophoresis the gel was stained with Coomassie brilliant blue, destained and processed for autoradiography. Lanes **a–c** represent Coomassie-stained lanes, lanes **d** and **e**, autoradiograms after 48 h exposure. The arrows mark crosslinked products.

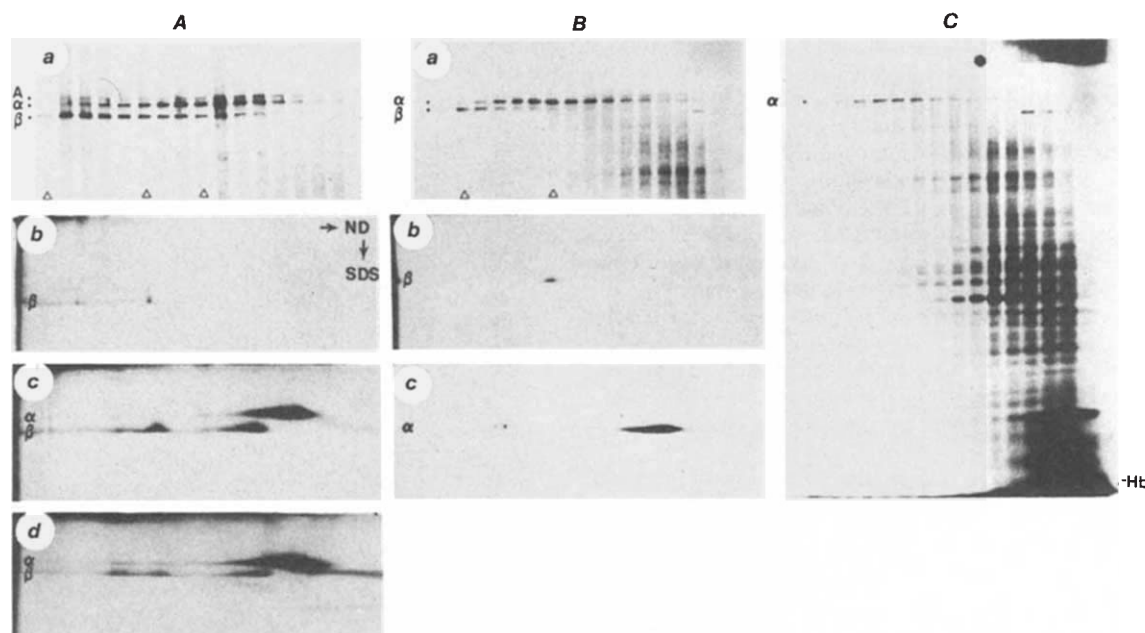


Fig. 3 Products of *in vitro* translation from pre-existing polyribosomes isolated from erythrocytes from 9-day chick embryos. Autoradiograms of the elution profiles of the soluble products of *in vitro* translation carried out in the absence (Aa) or presence of $40 \mu\text{g ml}^{-1}$ of rabbit erythrocyte spectrin-depleted, inside-out vesicle (IOVs; Ba) or $330 \mu\text{g ml}^{-1}$ of rabbit IOVs (C). Autoradiograms of the two-dimensional non-denaturing, SDS-PAGE analysis of the fractions, designated by Δ under the elution profiles, are depicted below their corresponding elution profiles in sequence. The smear running diagonally to the right-hand side from the main α -spectrin spot (most prominent in Ad) seems to represent nascent α -spectrin chains.

Methods. Erythrocytes from 9-day chick embryos were collected into sterile choline chloride (155 mM), HEPES (5 mM, pH 7.2) buffer on ice, concentrated by centrifugation and washed through sterile cotton wool to remove any contaminating white blood cells and platelets with 135 mM NaCl, 5 mM KCl, 10 mM HEPES (pH 7.2), 10 mM glucose and recentrifuged. Cells were then lysed in 1.8 vol of sterile distilled H_2O on ice and after 10 min, with periodic gentle mixing, were spun for 10 min at $13,000g$ in an Eppendorf centrifuge. The resultant supernatant was primed for *in vitro* translation as described for rabbit reticulocyte lysates³⁸ together with $300 \mu\text{Ci } ^{35}\text{S}$ -methionine per $475 \mu\text{l}$ of primed lysate. After incubation at 30°C for 45 min, rabbit spectrin-depleted IOVs were added at 0 (A), 40 (B) or $330 \mu\text{g ml}^{-1}$ (C). After a further 30 min of incubation at 30°C , the translation mixtures were centrifuged for 30 min at $150,000g$ in a Beckman airfuge with the vesicles pelleted through a 0.6 M sucrose pad. Aliquots ($250 \mu\text{l}$) of the supernatants were applied to a $1 \times 17 \text{ cm}$ Ac22 gel filtration column and fractions of $200 \mu\text{l}$ collected. Alternative fractions after the void volume were analysed by SDS-PAGE followed by autoradiography as described for Fig. 1.

The fractions designated by Δ were further analysed by two-dimensional non-denaturing followed by SDS-PAGE as described for Fig. 2.

Woods and Lazarides, personal observations). These observations, together with the above results, suggest that the fractions that assemble must exist in an alternative form to the homooligomeric conformations that accumulate in the unassembled pools. To approach this issue, we uncoupled synthesis from assembly onto the membrane and analysed any intermediates formed during assembly by a series of *in vitro* polysomal run-off experiments using hypotonic lysates made from erythroid cells isolated from 9–10-day chick embryos³¹. The Ac22 Ultrogel elution profile of the proteins synthesized *in vitro* was similar to that obtained with the *in vivo* labelled soluble proteins (Fig. 3A). In this case, however, there was a greater overlap of the β -spectrin with the α -spectrin peak, while a small subset of both proteins also eluted as monomers. Analysis of different fractions across the column by two-dimensional PAGE revealed that whereas there was no evidence for spectrin heterodimers in the soluble pool after *in vivo* labelling, close to 50% of the β -spectrin synthesized *in vitro* existed in the form of heterodimers with α -spectrin; the remainder of β -spectrin existed predominantly as homotetramers (Fig. 3A). Similarly, a fraction of the α -spectrin synthesized *in vitro* existed as heterodimers with β -spectrin, with the remainder existing mainly as homodimers (Fig. 3A). The proportion of β -spectrin in the form of heterodimers was remarkably close to that of newly synthesized β -spectrin which assembles on the membrane skeleton *in vivo*²⁵.

The heterodimeric spectrin produced *in vitro* bound to rabbit erythrocyte spectrin-depleted IOVs but not to right-side-out vesicles, ankyrin-depleted or acetic acid-stripped IOVs, similarly

to the homotetrameric β -spectrin (Table 1c). Binding studies at different IOV concentrations revealed preferential binding of the heterodimeric over the β -homotetrameric species at sub-saturating concentrations of IOVs. At a concentration of IOVs equivalent to $330 \mu\text{g}$ membrane protein per ml, both β -spectrin-containing forms bound effectively to the IOVs (Fig. 3C, Table 1c). With $40 \mu\text{g}$ membrane protein per ml, all of the heterodimeric spectrin bound to the membrane whereas most of the β -homotetramer remained in the soluble fraction (Fig. 3B, Table 1c). Such experiments suggest that these heterodimers constitute the fraction that rapidly assembles on the membrane *in vivo*, and that their assembly occurs too rapidly for detection after *in vivo* labelling of erythroid cells from 10-day-old chick embryos.

In conclusion, the data presented here indicate that newly synthesized α - and β -spectrin monomers rapidly assume one of two configurations, an $\alpha\beta$ -heterodimer or a homo-oligomer (β_4 or α_2). Interestingly, only newly synthesized spectrin polypeptides can assume a stable homo-oligomeric conformation, suggesting that the folding of the nascent polypeptide may begin from the NH_2 - towards the COOH -terminus in parallel with synthesis. The diffusion-collision theory of protein-folding dynamics predicts that native secondary structure is achieved within each microdomain of a polypeptide by random events and is stabilized by subsequent interactions between microdomains^{32,33}. Human α - and β -spectrin polypeptides are known to consist of multiple homologous non-identical repeating triple helical segments, each 106 amino acids long^{4,7}, as has also been observed in chicken α -spectrin^{34,35}. If this model is correct, these

Table 1 Membrane-binding ability of unassembled β - and α -spectrin

a, Partially purified by gel filtration				
Membranes added	β -Spectrin	% Pelleted		Mixture of
		α -Spectrin	$\alpha + \beta$ -spectrin	α β
0	5	5	5	6
ROVs	14	11	19	11
IOVs	97	10	18	92
Acetic-acid stripped				
IOVs	16	22	21	15
Trypsinized IOVs	8	12	10	3
b, Total lysate				
Membranes	β -Spectrin	α -Spectrin		
0	9	9		
ROVs	11	8		
IOVs	97	19		
Acetic-acid stripped				
IOVs	12	11		
c, Products of <i>in vitro</i> translation				
Addition of IOVs				
($\mu\text{g ml}^{-1}$ membrane protein)				
330	97	46		
124	93	42		
40	52	41		
4	31	27		
0	14	16		

Aliquots of the peak fractions of pulse-labelled β -spectrin and α -spectrin, partially purified by gel-filtration chromatography as described in Fig. 1A legend. (a), total soluble hypotonic lysate, prepared immediately after a 10-min pulse of erythroid cells from 11-day chick embryos (b) or the products of *in vitro* translation as described for Fig. 4 (c) were incubated either alone or in the presence of rabbit erythrocyte ROVs, spectrin-depleted IOVs, acetic-acid-stripped, ankyrin-depleted IOVs or trypsinized IOVs prepared by the protocol described by Bennett and Branton³⁷. Equal aliquots of ghost erythrocyte membranes were used for the preparation of each of these types of vesicles and the resultant vesicles suspended in identical volumes. The volumes added to the soluble spectrin fraction gave concentrations of IOV membrane protein equivalents of $330 \mu\text{g ml}^{-1}$. After a 20-min incubation at 38°C , the vesicles were pelleted through a 0.6 M sucrose cushion in Tris-HCl (5 mM, pH 7.2) for 20 min at $150,000g$ in a Beckman airfuge. The relative amounts of α - and β -spectrin in the supernatants and pellets were analysed by SDS-PAGE followed by autoradiography and densitometry of the resultant autoradiograms. The values represent the mean values from three separate experiments; variability was less than 7%.

helices will presumably begin to form as each 106-amino-acid segment is synthesized. The fact that the two spectrin molecules are so long ($\sim 4,000$ amino acids each) probably increases the chances of coiled-coil interactions occurring between adjacent nascent chains before synthesis is complete. Therefore, the formation of stable spectrin homo-oligomeric species *in vivo* may result from the greater chance of coiled-coil interactions occurring between like rather than unlike spectrin subunits, even though the $\alpha\beta$ form may be thermodynamically favoured.

It has been proposed (ref. 25) that spectrin and ankyrin assembly occurs through a cascade of limiting steps, where the extent of α -spectrin assembled is limited by the amount of β -spectrin assembled, which in turn is limited by the amount of ankyrin bound. The data presented here suggest instead that the limiting factor is the extent of heterodimerization. Furthermore, homo-oligomerization prevents the unassembled spectrins from reacting with each other or with ankyrin and in the case of α -spectrin also from binding to the plasma membrane. Although β -homotetramers can bind to spectrin-depleted plasma-membrane vesicles, presumably to ankyrin, they appear to bind with lower affinity than heterodimers. The difference in

membrane affinity, coupled with the fact that unassembled β -spectrin (β_4 -spectrin) is degraded extremely rapidly and selectively²⁹, makes it feasible that if any β -homotetramers do bind to the membrane *in vivo*, they are displaced by $\alpha\beta$ -heterodimers and then degraded rapidly, thereby ensuring the correct stoichiometry of assembled spectrin.

Finally, it is interesting to note that the β -homotetramer appears to bind to preassembled ankyrin, yet no association can be detected with the unassembled ankyrin that accumulates transiently in the soluble pool. This paradox suggests that this unassembled ankyrin must also exist in some modified form unable to interact with β -spectrin.

We thank Drs Horst Hinssen, John Cox and John Ngai for their valuable comments on the manuscript. This work was supported by grants from the NIH, NSF and the Muscular Dystrophy Association. C.M.W. was also supported by a grant from the Camille and Henry Dreyfus Foundation.

Received 20 September 1985; accepted 4 February 1986.

1. Branton, D., Cohen, C. M. & Tyler, J. *Cell* **24**, 24–32 (1981).
2. Sheetz, M. P. *Semin. Hemat.* **20**, 175–188 (1983).
3. Yu, J., Fischman, D. A. & Steck, T. J. *supramolec. Struct.* **1**, 233–248 (1973).
4. Speicher, D. & Marchesi, V. T. *Nature* **311**, 177–180 (1984).
5. Shotton, D., Burke, B. & Branton, D. *Biochim. biophys. Acta* **536**, 303–329 (1978).
6. Cohen, C. M. *Semin Hemat.* **20**, 141–158 (1983).
7. Marchesi, V. T. A. *Rev. Cell Biol.* (in the press).
8. Ungewickell, E., Bennett, P., Calvert, R., Ohanian, V. & Gratzner, W. B. *Nature* **280**, 811–814 (1979).
9. Fowler, V. & Taylor, D. L. *J. Cell Biol.* **85**, 361–376 (1980).
10. Pinder, J. C. & Gratzner, W. B. *J. Cell Biol.* **96**, 768–775 (1983).
11. Bennett, V. & Stenbuck, P. J. *J. biol. Chem.* **254**, 2533–2541 (1979).
12. Luna, E. J., Kidd, G. H. & Branton, D. *J. biol. Chem.* **254**, 2526–2532 (1979).
13. Calvert, R., Bennett, P. & Gratzner, W. B. *Eur. J. Biochem.* **107**, 355–361 (1980).
14. Litman, D., Hsu, C. J. & Marchesi, V. T. *J. Cell Sci.* **42**, 1–22 (1980).
15. Bennett, V. & Stenbuck, P. *Nature* **280**, 468–473 (1979).
16. Bennett, V. & Stenbuck, P. *J. biol. Chem.* **255**, 6424–6432 (1980).
17. Hargreaves, W. R., Giedd, K. N., Verkleij, A. & Branton, D. *J. biol. Chem.* **255**, 11965–11972 (1980).
18. Anderson, R. A. & Lovrien, R. E. *Nature* **307**, 655–658 (1984).
19. Pasternak, G. R., Anderson, R. A., Leto, T. L. & Marchesi, V. T. *J. biol. Chem.* **260**, 3676–3683 (1985).
20. Yoshino, H. & Marchesi, V. T. *J. biol. Chem.* **259**, 4496–4500 (1984).
21. Ungewickell, E. & Gratzner, W. B. *Eur. J. Biochem.* **88**, 379–385 (1978).
22. Morrow, J. S. & Marchesi, V. T. *J. Cell Biol.* **88**, 463–468 (1981).
23. Liu, S.-C., Windisch, P., Kim, S. & Palek, J. *Cell* **37**, 587–594 (1984).
24. Blikstad, I., Nelson, W. J., Moon, R. T. & Lazarides, E. *Cell* **32**, 1081–1091 (1983).
25. Moon, R. T. & Lazarides, E. *J. Cell Biol.* **98**, 1899–1904 (1984).
26. Cox, J. V., Moon, R. T. & Lazarides, E. *J. Cell Biol.* **100**, 1548–1557 (1985).
27. Granger, B. L. & Lazarides, E. *Nature* **313**, 238–241 (1984).
28. Weise, M. J. & Chan, L.-N. L. *J. biol. Chem.* **253**, 1892–1897 (1978).
29. Woods, C. M. & Lazarides, E. *Cell* **40**, 959–969 (1985).
30. Davies, J. & Bennett, V. *J. biol. Chem.* **258**, 7757–7766 (1983).
31. Moon, R. T. & Lazarides, E. *Nature* **305**, 62–64 (1983).
32. Karplus, M. & Weaver, D. L. *Nature* **260**, 404–406 (1976).
33. Harrison, S. C. & Durbin, R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4028–4030 (1985).
34. Wasenius, V.-H., Saraste, M., Knowles, J., Virtanen, I. & Lehto, V.-P. *EMBO J.* **4**, 1425–1430 (1985).
35. Birkenmeier, C. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5671–5675 (1985).
36. Hubbard, B. & Lazarides, E. *J. Cell Biol.* **80**, 166–182 (1979).
37. Bennett, V. & Branton, D. *J. biol. Chem.* **252**, 2753–2763 (1977).
38. Jackson, R. T. & Hunt, T. *Meth. Enzym.* **96**, 50–73 (1983).

Template-free RNA synthesis by Q β replicase

Christof K. Biebricher, Manfred Eigen & Rüdiger Luce

Max-Planck-Institut für Biophysikalische Chemie,
D-3400 Göttingen, FRG

In the absence of extraneously added template, standard preparations of Q β replicase spontaneously synthesize RNA *in vitro*, possibly as a result of RNA contamination¹. Using special enzyme purifications, Sumper and Luce² presented evidence that self-replicating RNA not present *ab initio* can grow out of 'template-free' incorporation mixtures. In contrast to DNA polymerase I and RNA polymerase, which also show *de novo* synthesis^{3–5}, the products synthesized 'de novo' by Q β replicase are RNA species containing nonrepetitive sequences of defined lengths which differ between experiments, even when synthesized under identical condi-

tions, in fingerprints, chain lengths and kinetic parameters⁶. Kinetic analysis of the *de novo* processes⁷ distinguished it from template-instructed synthesis and excluded an assumption of self-replicating RNA contamination. These conclusions were questioned recently by Hill and Blumenthal⁸, who claimed to show that highly purified Q β replicase preparations cannot produce RNA *de novo*. We now present evidence that, under the conditions required for *de novo* synthesis, Q β replicase prepared according to their method is also capable of *de novo* synthesis. Furthermore, we show that Q β replicase condenses nucleoside triphosphates to more or less random oligonucleotides.

Hill and Blumenthal⁸ presented kinetic profiles of RNA synthesis for standard Q β replicase preparations, for which RNA synthesis was detectable after about 30 min, and for specially purified R(-S1) Q β replicase, for which no RNA synthesis was detected over a 4 h; addition of heat-treated standard enzyme preparations restored nucleotide incorporation in a way they claimed to be "kinetically indistinguishable from *de novo* synthesis". None of their data show the characteristics of what we called *de novo* synthesis. Instead, their Figs 1 and 3 identify the underlying process as template-instructed, with much shorter lag times than necessary for amplification of a single strand of self-replicating RNA to macroscopic appearance (see Table IV of ref. 7). Their finding that "the length of the lag period" was "inversely proportional to the concentration of nucleoside triphosphates" is also typical of template-instructed reactions, quite in contrast to the process we called *de novo* synthesis, which depends on the 2.8th power of nucleotide and 2.5th power of enzyme concentrations. (Both exponents represent lower bounds.) Hence, their incorporation products indeed did not result from *de novo* synthesis but rather from template-instructed amplification of RNA impurities present in their standard enzyme preparations.

Using enzyme purified by a different method, they detected no incorporation up to 4 h without added template. At R(-S1) enzyme concentrations as low as used in their experiments (68 nM), we also observed no *de novo* synthesis⁷ over this time. Other groups have performed experiments with Q β replicase⁹⁻¹¹ without reporting interference by *de novo* synthesis, and examination of the conditions used by these groups shows in each case that no *de novo* synthesis would be expected.

Unfortunately, Hill and Blumenthal were unable to cooperate in an effort to determine whether their purification procedure yields an enzyme unable to synthesize RNA *de novo* (T. Blumenthal, personal communication). We therefore repeated their experiments independently as well as was possible using the sparse data given in their report. Our results are:

(1) Enzyme prepared according to Blumenthal's standard procedure¹², despite being electrophoretically ~90% pure, contained RNA at optically detectable levels, part of which was self-replicating as macroscopic RNA synthesis started after short lag times. The products were defined RNA species reproducibly selected under identical conditions. However, even minor changes in conditions often resulted in different electrophoretic patterns of the products.

(2) Although phosphocellulose-urea purification step introduced by Hill and Blumenthal indeed removed the RNA, the acidic enzyme subunits S1, and later elongation factors (EFs) Tu and Ts, eluted slowly from the column under their conditions. Depending on the wash volume and time of the urea step, only 10–50% of the input viral β -subunit was recovered. Resulting preparations still contained 10% S1, but excess EF Tu and Ts.

(3) Excess EF Tu, Ts and residual holoenzyme could be removed in the next step if the described stepwise elution was replaced by gradient elution^{2,13}. SDS-gel electrophoresis combined with silver staining revealed no impurities in the resulting preparation or in our standard preparations (see Fig. 3 in ref. 14). However, the low synthesis rate found by Hill and Blumenthal ($1.6 \times 10^{-3} \text{ s}^{-1}$ compared with our $3.3\text{--}6.5 \times 10^{-3} \text{ s}^{-1}$) led us to suspect that the concentration of active enzyme in their

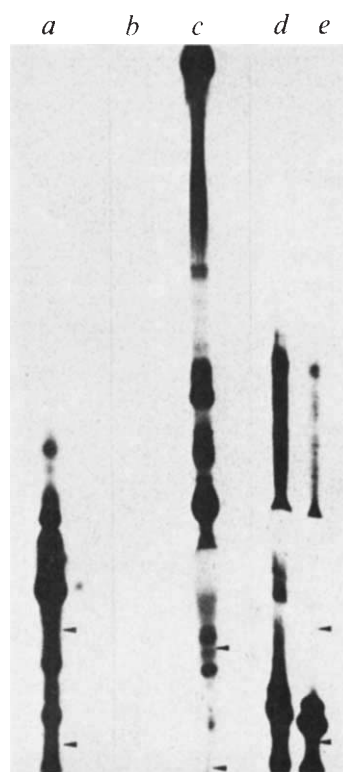


Fig. 1 Electropherograms showing condensation of triphosphates to oligonucleotides by Q β replicase. The incorporation mixtures contained in Q β buffer (0.05 M Tris-HCl buffer pH 7.5, 0.2 mM dithiothreitol, 10 mM MgCl₂, 20% glycerol) were: lane a, 500 μ M [α -³²P]ATP, 200 μ M GTP, 50 μ M PP_i, 500 nM replicase; lanes b, c, 200 μ M GTP, 50 μ M ATP, 1.5 μ M [α -³²P]CTP, 20 μ M PP_i, 200 nM replicase; lane d, 800 μ M ATP, 800 μ M GTP, 20 μ M PP_i, 600 nM replicase; lane e, 800 μ M ITP, 800 μ M GTP, 20 μ M PP_i, 600 nM replicase.

Methods. Incubations were for 5 min (b), 68 h (a, c) or 16 h (d, e) at 30 °C. The oligonucleotides from incubations d and e were treated with phosphatase, isolated on QAE-Sephadex-A25 and labelled with polynucleotide kinase and [γ -³²P]ATP before loading onto the gel. The incubation mixtures were analysed on sequencing gels (12% acrylamide). The arrows give the positions of the dyes bromophenol blue (mobility of the undecamer) and dinitrophenol (mobility of the pentamer). Electrophoresis was from top to bottom. The material eluted from the gels was degradable by alkali or snake venom phosphodiesterase.

preparation was overestimated.

(4) Using conditions appropriate for observing *de novo* synthesis, we found that replicase purified by their method has about the same ability to synthesize RNA *de novo* as was observed with our preparations. The essential point is to choose sufficiently high enzyme (and nucleoside triphosphate) concentrations to utilize the strong concentration dependence of lag times (see Fig. 4 of ref. 7). Concentration of the enzyme on QAE-Sephadex (step VIII of ref. 2) was therefore required for the *de novo* experiments. In accordance with earlier reports⁶, the products were different even when synthesized under identical conditions.

(5) Enzyme preparations isolated by the standard purification method² from uninfected *Escherichia coli* containing a plasmid carrying the structural gene of the viral replicase showed the same enzyme concentration threshold for *de novo* synthesis and the same product characteristics as found with preparations from Q β -infected bacteria.

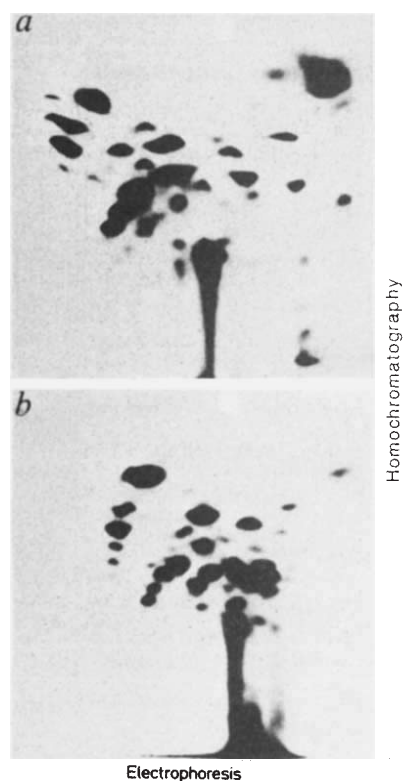
We conclude that the experiments of Hill and Blumenthal do not support their interpretation. Our original kinetic analysis was, as reported⁷, an adequate method to disprove a template-instructed process resulting from RNA impurities. Although the molecular mechanism behind this process is unknown, a

Fig. 2 Fingerprinting of the oligonucleotides. Incorporation mixture (100 μ l): 200 μ M GTP, 200 μ M ATP, 200 μ M UTP, 10 mM $MgCl_2$, 1 mM $MnCl_2$, 400 nM holoreplicase in Q β -buffer. *a*, No addition; *b*, 400 nM yeast 5.8 S RNA.

Methods. The mixture was incubated for 21 h at 30 °C. The enzyme was denatured for 1 min at 100 °C and removed by centrifugation. The solution was diluted with 500 μ l 0.02 M Tris base and treated with 8 U alkaline phosphatase (calf intestine, Boehringer, Mannheim) for 1 h at 37 °C. The products were isolated by loading the mixtures on 1-ml columns of QAE-Sephadex-A25 Cl^- form. The column was washed with 0.05 M Tris buffer (pH 7.5) and the oligonucleotides eluted with 0.5 M NaCl, desalted on BioGel P2, lyophilized twice and dissolved in 20 μ l H_2O . The yields were 0.08 A_{260} (80 μ M nucleotide incorporated) in the absence of RNA and 0.21 A_{260} (210 μ M nucleotide, 35% incorporated) in the presence of yeast 5.8 S RNA. 500 pmol oligonucleotides were labelled at the 5' ends with [γ - ^{32}P]ATP and polynucleotide kinase (1 h, 37 °C), isolated on BioGel P2 and fingerprinted according to ref. 18. Analysis of the 5' end of the oligonucleotides: The incorporations were with 5.8 S RNA as described above. Yield of 5' ends, 43 μ M: the 5'-labelled oligonucleotide material was degraded with 10% piperidine and the products separated on polyethyleneimine-impregnated thin layer plates. *pGp was found to be the only labelled product and its radioactivity was measured. Incorporation of [γ - ^{32}P]GTP, 48 μ M: incubation as in *a* except that [γ - ^{32}P]GTP was used. The material was separated on Biogel, degraded with piperidine followed by treatment with snake venom phosphodiesterase to destroy GTP. The degradation products were separated by electrophoresis on DEAE paper in 7% formic acid. The *pppGp spot was cut out and the yield measured by counting the radioactivity. Yield of pppGpA-ends, 29 μ M: analysis was as for [γ - ^{32}P]GTP, except that [α - ^{32}P]ATP was used and the pppGp* spot was counted. The degradation products also included Gp*, Ap*, *pA and Up*. The quantitation method was checked by measuring template-instructed incorporation of [γ - ^{32}P]GTP into MNV-11, isolating the products and analysing them as for [γ - ^{32}P]GTP.

hypothesis agreeing with the experimental data has been presented^{4,6,7}, and consists of two phases. In the first, polymerase slowly catalyses random synthesis of oligonucleotides, and in the second, once a suitable template is formed by chance, rapid evolution and autocatalytic amplification to macroscopic appearance occur. It is not surprising that the *de novo* products of DNA polymerase I and RNA polymerase are so different from those of Q β replicase. The slipping homo- or co-polymer products of DNA polymerase I are indeed the only templates it can amplify autocatalytically, whereas they cannot be amplified autocatalytically by Q β replicase, because RNA with high internal stabilization of its single-stranded structure is required to escape double-strand formation with replica¹⁵. It is doubtful whether the first phase can be realized at all with DNA polymerase I, which is unable to initiate a new chain, as the lag times suggest autocatalytic amplification of preexisting impurities^{3,16}. The latter possibility, however, is excluded by the kinetic analysis for Q β replicase⁷, whose two nucleotide binding sites would provide, in principle, for slow condensation of nucleotide triphosphates to oligonucleotides.

Under normal conditions, template-instructed nucleotide incorporation proceeds much too rapidly for observable accumulation of the initial *de novo* synthesis products⁶. We blocked RNA amplification by omitting one triphosphate, and were successful when one or both of the pyrimidine nucleotides were omitted; however, ATP and GTP were both required for nucleotide condensation to oligonucleotides. Pyrimidines, if present, were also incorporated. The rate of phosphodiester formation under these conditions was four or five orders of magnitude below that of the template-instructed process. The chain lengths of the oligonucleotides that accumulated to larger concentrations ranged from 5 to 20 (Fig. 1); the fingerprints suggest that there is little sequence preference (Fig. 2), although the 5' end was always pppGpR- (Fig. 2 legend). Elongation of preexisting oligonucleotides by Q β replicase therefore did not occur under these conditions. The presence of 5.8S yeast RNA,



but not transfer RNA, doubled the incorporation rate. The screening of oligonucleotides synthesized with 5.8 S RNA present showed no biases (Fig. 2), as would be expected if there was any instruction by the RNA.

For RNA synthesis in the absence of extraneously added template, we find: (1) The self-replicating RNA species produced in each template-free experiment differ from each other in sequence and chain length even in experiments performed under identical conditions. (2) Kinetic analysis of the *de novo* synthesis process⁷ reveals fundamental differences from the well-understood kinetics of template-instructed replication¹⁷. (3) When replication is inhibited by omitting one triphosphate, many different oligonucleotides accumulate. (4) Highly purified preparations of Q β replicase isolated from uninfected *E. coli* cells containing plasmids carrying the structural gene of the viral replicase show the same *de novo* synthesis characteristics as those isolated from infected bacteria.

No alternative has been presented to our original interpretation of these results, that Q β replicase, in the absence of self-replicating RNA templates, is able to synthesize them in a non-instructed way by trial and selection.

We thank Dr M. A. Billeter for supplying us with the bacterial strain carrying a plasmid with the replicase gene of Q β , and Dr W. C. Gardiner for critical reading of the manuscript.

Received 20 February; accepted 5 March 1986.

1. Weissmann, C., Billeter, M. A., Goodman, H. M., Hindley, J. & Weber, H. A. *Rev. Biochem.* **42**, 303-328 (1973).
2. Sumper, M. & Luce, R. *Proc. natn. Acad. Sci. U.S.A.* **72**, 162-166 (1975).
3. Radding, C. R. & Kornberg, A. *J. biol. Chem.* **237**, 2877-2882 (1962).
4. Kornberg, A., Bertsch, L. L., Jackson, J. F. & Khorana, H. G. *Proc. natn. Acad. Sci. U.S.A.* **51**, 315-323 (1964).
5. Krakow, J. S. & Karstadt, M. *Proc. natn. Acad. Sci. U.S.A.* **58**, 2094-2101 (1967).
6. Biebricher, C. K., Eigen, M. & Luce, R. *J. molec. Biol.* **148**, 369-390 (1981).
7. Biebricher, C. K., Eigen, M. & Luce, R. *J. molec. Biol.* **148**, 391-410 (1981).
8. Hill, D. & Blumenthal, T. *Nature* **301**, 350-352 (1983).
9. Palmenberg, A. & Kaesberg, P. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1371-1375 (1974).
10. Kramer, F. R., Mills, D. R., Cole, P. E., Nishihara, T. & Spiegelman, S. *J. molec. Biol.* **89**, 719-736 (1974).
11. Mills, D. R., Peterson, R. L. & Spiegelman, S. *Proc. natn. Acad. Sci. U.S.A.* **58**, 217-224 (1967).
12. Blumenthal, T. *Meth. Enzym.* **60**, 628-638 (1979).
13. Cole, P. E., Sawchyn, I. & Guerrier-Takada, C. *J. biol. Chem.* **257**, 12929-12934 (1982).
14. Biebricher, C. K. *Evol. Biol.* **16**, 1-52 (1983).
15. Biebricher, C. K., Eigen, M. & Gardiner, W. C. *Biochemistry* **23**, 3186-3194 (1984).
16. Kornberg, A. *DNA Replication* (Freeman, San Francisco 1980).
17. Biebricher, C. K., Eigen, M. & Gardiner, W. C. *Biochemistry* **22**, 2544-2559 (1983).
18. Volckaert, G., Min Jou, W. & Fiers, W. *Analyt. Biochem.* **72**, 433-446 (1976).

Applied Biological Sciences

Use of Small Fish Species in Carcinogenicity Testing. Monograph 65. KAREN L. HOOVER (ed.). U.S. Department of Health and Human Services:1984. Pp.409. Np.

Veterinary Immunology. By P.M. OUTERIDGE. Academic:1985. Pp.280. ISBN 0-12-531130-3. \$45, £36.

Veterinary Microbiology: Molecular and Clinical Perspectives. R. PANDEY (ed.). Karger:1985. Pp.222. Hbk ISBN 3-8055-4067-1. SwFr.148, DM177, £46.30, \$63.

Sociology

Circulation in Population Movement: Substance and Concepts from the Melanesian Case. By MURRAY CHAPMAN and R. MANSELL PROTHERO. Routledge & Kegan Paul:1985. Pp.483. ISBN 0-7102-045105. £30.

Home Environments: Human Behaviour and Environment. IRWIN ALTMAN and CAROL M. WERNER (eds). Plenum:1985. Pp.339. ISBN 0-306-41976-9. £39.50.

Reproductive Change in Developing Countries: Insights from the World Fertility Survey. JOHN CLELAND and JOHN HOBcraft (eds). Oxford University Press:1985. Pp.301. ISBN 0-19-828465-9. £15.

Archaeology

The Cambridge Field Guide to Prehistoric Life. By DAVID LAMBERT. Cambridge University Press:1985. Pp.256. ISBN 0-521-26685-8. Hbk £17.50; pbk £6.95.

Contributions to the Archaeology and Ethnohistory of Greater Mesoamerica. WILLIAM J. FOLAN (ed.). Southern Illinois University Press:1985. Pp.334. ISBN 0-8093-1123-2. \$28.95.

General

The Cape Cod National Seashore: A Landmark Alliance. By CHARLES H.W. FOSTER. University Press of New England:1985. Pp.125. Pbk ISBN 0-87451-346-4. Pbk £8.60, \$10.75.

Do You Think . . . By HAROLD VAN COLLE. Harold van Colle, 9 Mayfields Close, Wembley Park, Middlesex HA9 9PP, UK:1985. Pp.54. Pbk £3.

European Brewery Convention. Proceedings of the 20th Congress, Helsinki 1985. IRL Press:1985. Pp.784. Hbk ISBN 0-947946-42-X. £50, \$90.

Forest Products: Advanced Technologies and Economic Analyses. By DAVID A. TILLMAN. Academic:1985. Pp.283. ISBN 0-12-691270-X. \$36, £36.

The Geography of Peace and War. DAVID PEPPER and ALAN JENKINS (eds). Basil Blackwell:1985. Pp.222. ISBN 0-631-13559-6. Hbk £24.50; pbk £7.95.

Handbook of Discourse Analysis, Vol.2: Dimensions of Discourse. TEUN A. VAN DIJK (ed.). Academic:1985. Pp.279. ISBN 0-12-712002-5. \$53.50, £45.50.

Histoire de ma Jeunesse. By FRANCOIS ARAGO. Christian Bourgois Editeur:1985. Pp.189. Pbk ISBN 2-267-00426-7. Np.

A History of British Atomic Tests in Australia. By J.L. SYMONDS. Australian Government Publishing Service:1985. Pp.593. ISBN 0-644-04118-8. Np.

ICSU Short Reports, Vol. 3: Third European Bioenergetics Conference. GÜNTER SCHÄFER (ed.). Cambridge University Press:1985. Pp.745. ISBN 0-521-30813-5. £50.

The Ins and Outs of Peg Solitaire. By JOHN D. BEASLEY. Oxford University Press:1985. Pp.275. ISBN 0-19-853203-2. £12.50.

The International Primate Trade, Vol. 1. DAVID MACK and RUSSELL A. MITTERMEIER (eds). World Wildlife Fund:1985. Pp.185. Pbk \$20.

Islands of History. By MARSHALL SAHLINS. University of Chicago Press:1985. Pp.180. ISBN 0-226-73357-2. £21.50.

Just Health Care. By NORMAN DANIELS. Cambridge University Press:1985. Pp.245. ISBN 0-521-23608-8. Np.

Language Learnability and Language Development. By STEVEN PINKER. Harvard University Press:1985. Pp.435. ISBN 0-674-51054-2. \$29.50, £26.95.

Learn to Paint Wildlife. By MARTIN KNOWLEDEN. Collins:1985. Pp.64. Pbk ISBN 0-00-411970-3. Pbk £4.95.

Living with the Bomb. Can We Live Without Enemies? By DOROTHY ROWE. Routledge & Kegan Paul:1985. Pp.243. Pbk ISBN 0-7102-0477-9. £4.95.

Machine Independent Organic Software Tools (MINT), 2nd Edn. By M.D. GODFREY. Academic:1985. Pp.375. ISBN 0-12-286983-4. \$30, £25.

MCDM: Past Decade and Future Trends: A Source Book of Multiple Criteria Decision Making. MILAN ZELENY (ed.). JAI Press:1985. Pp.337. ISBN 0-89232-439-2. \$57.50.

Medical Rehabilitation. LAURO S. HALSTEAD and MARTIN GRABOIS (eds). Raven Press:1985. Pp.354. ISBN 0-88167-085-5. \$65.

Methodes de Datation par les Phenomenes Nucleaires Naturels, Applications. ETIENNE ROTH, BERNARD POTY and MARIE THÉRESE MENAGER. Masson:1985. Pp.631. Hbk ISBN 2-225-80674-8. Np.

Metropolitan Desegregation. ROBERT L. GREEN (ed.). Plenum:1985. Pp.234. ISBN 0-306-41964-5. Np.

Money into Light: The Emerald Forest, A Diary. By JOHN BOORMAN. Faber:1985. Pp.229. Pbk ISBN 0-571-13772-5. Pbk £4.95.

Nemesis: The Death-Star and Other Theories of Mass Extinction. By DONALD GOLDSMITH. Walker:1985. Pp.166. ISBN 0-8027-0872-2. \$15.95.

New Methods and Techniques for Publishers and Learned Societies. M. FEENEY (ed.). Primary Communications Research Centre, University of Leicester:1985. Pp.288. Pbk ISBN 0-906-083-29-X. Np.

The Nuclear Confrontation in Europe. JEFFREY D. BOUTWELL, PAUL DOTY and GREGORY F. TREVERTON (eds). Croom Helm:1985. Pp.247. ISBN 0-7099-2481-X. £22.50.

On the Frontiers of Science: Strange Machines You Can Build. By G. HARRY STINE. Atheneum:1985. Pp.111. Pbk ISBN 0-689-11562-8. Pbk \$8.95.

Operating Systems: Structures and Mechanisms. By PHILIPPE A. JANSON. Academic:1985. Pp.267. ISBN 0-12-380230-X. \$29.50, £19.50.

Passing Farms: Enduring Values. California's Santa Clara Valley. By YVONNE JACOBSON. William Kaufmann:1984. Pp.240. ISBN 0-86576-045-4. \$39.50.

Pharmacological and Chemical Synonyms, 8th edn. Compiled by E.E.J. MARLER. Elsevier:1985. Pp.545. Hbk ISBN 0-444-90359-3. Dfl.245, \$94.25.

Philosophy Through the Looking Glass: Language, Nonsense, Desire. By JEAN JACQUES LECERCLE. Hutchinson:1985. Pp.206. Pbk ISBN 0-09-161371-X. £5.95.

The Politics of Nuclear Defense: A Comprehensive Introduction. By GREVILLE RUMBLE. Polity Press, Cambridge, UK:1985. Pp.285. Hbk ISBN 0-7456-0194-4; pbk ISBN 0-7456-0195-2. Hbk £19.50; pbk £6.95.

The Pond. By GERALD THOMPSON, JENNIFER COLDREY and GEORGE BERNARD. Collins:1985. Pp.256. Pbk ISBN 0-00-219287-X. Pbk £9.95.

Principles of Phase Conjugation. By B.Y. ZEL'DOVICH, N.R. PILIPETSKY and V.V. SHKUNOV. Springer-Verlag:1985. Pp.250. ISBN 3-540-13458-1/0-387-13458-1. DM 140.

Quality Control in the Food Industry, Vol. 1, 2nd Edn. S.M. HERSCHDOERFER (ed.). Academic:1985. Pp.469. ISBN 0-12-343001-1. \$45, £45.

Random Graphs. By BELLA BOLLOBAS. Academic:1985. Pp.447. Hbk ISBN 0-12-111755-3. £52, \$58.50.

Reflexivity: The Post Modern Predicament. By HILARY LAWSON. Hutchinson:1985. Pp.132. Pbk ISBN 0-09-160-861-9. £5.50.

Risk Analysis in the Private Sector. CHRIS WHIPPLE and VINCENT T. COVELLO (eds). Plenum:1985. Pp.506. ISBN 0-306-41924-6. Np.

Risk Quantification and Regulatory Policy. Banbury Report 19. DAVID G. HOEL, RICHARD A. MERILL and FEDERICA P. PERERA (eds). Cold Spring Harbor Laboratory, New York, USA:1985. Pp.368. Hbk ISBN 0-87969-219-7. \$67.

Science and Technology for International Development: An Assessment of U.S. Policies and Programs. By ROBERT P. MORGAN. Westview Press:1985. Pp.160. ISBN 0-86531-838-7. £17.50.

The Second Self: Computers and the Human Spirit. By SHERRY TURKLE. Simon & Schuster:1985. Pp.362. Pbk ISBN 0-671-60602-6. Pbk \$8.95.

Seven Ideas That Shook the Universe. By NATHAN SPIELBERG and BRYON D. ANDERSON. Wiley:1985. Pp.291. ISBN 0-471-81477-6. £19.95.

Short Term Bioassays in the Analysis of Complex Environmental Mixtures IV. MICHAEL D. WATER et al. (eds). Plenum:1985. Pp.384. ISBN 0-306-42015-5. Np.

Should the Baby Live? The Problem of Handicapped Infants. By HELGA KUHSE and PETER SINGER. Oxford University Press:1985. Pp.228. Hbk ISBN 0-19-217745-1; pbk ISBN 0-19-286062-3. Hbk £12.50; pbk £3.95.

Space Law in the United Nations. By MARIETTA BENKO et al. Martinus Nijhoff:1985. Pp.256. ISBN 90-247-3157-7. Dfl. 115, \$38.

Statistics for Land Managers. By FRANK FREESE. Paeony Press, Allerley Brae, Oxnam Road, Jedburgh, Roxburghshire TD8 6JL, UK:1984. Pp.176. ISBN 0-946941-00-9. £6.50.

Striking a Balance: National Security and Scientific Freedom. First Discussions. HAROLD C. RELYEA (ed.). American Association for the Advancement of Science, 1333 H Street, NW, Washington DC 20005, USA:1985. Pp.133. Np.

Supersymmetry. K. DIETZ et al. (eds). Plenum:1985. Pp.708. ISBN 0-306-42012-0. Np.

Surveying Subjective Phenomena, Vol. 1. CHARLES F. TURNER and ELIZABETH MARTIN (eds). Russell Sage Foundation, 112 East 64th Street, New York:1985. Pp.495. Hbk ISBN 0-87154-882-8. \$37.50.

Surveying Subjective Phenomena, Vol. 2. CHARLES F. TURNER and ELIZABETH MARTIN (eds). Russell Sage Foundation, 112 East 64th Street, New York:1985. Pp.617. Hbk ISBN 0-87154-883-6. \$47.50.

Technology, Economic Growth and the Labour Process. By PHIL BLACKBURN, ROD COOMBS and KENNETH GREEN. Macmillan Press:1985. Pp.239. ISBN 0-333-37496-7. £25.

To Whom It May Concern: An Investigation of the Art of Elephants. By DAVID GUCWA and JAMES EHMANN. W.W. Norton:1985. Pp.269. Hbk ISBN 0-393-02240-4. \$14.95.

The Travelling Salesman Problem: A Guided Tour of Combinatorial Optimization. E.L. LAWLER et al. (eds). Wiley:1985. Pp.465. ISBN 0-471-90413-9. £39.95.

Twenty Years Before the Mast. By CHARLES ERSKINE. Smithsonian Institution Press:1985. Pp.311. Pbk ISBN 0-87474-415-6. \$4.95.

Understanding Language: Man or Machine. By JOHN A. MOYNE. Plenum:1985. Pp.357. ISBN 0-306-41970-X. \$49.50.

The Urban Edge: Where the City Meets the Sea. JOSEPH E. PETRILLO and PETER GRENELL (eds). William Kaufmann:1985. Pp.108. ISBN 0-86576-078-0. Np.

Water Law. By WILLIAM GOLDFARB. Ann Arbor Science:1985. Pp.233. Hbk ISBN 0-250-40627-6. £48.00.

The Wildlife Parks of Africa. By NICHOLAS LUARD. Michael Joseph:1985. Pp.240. Hbk ISBN 0-7181-2485-5. Hbk £14.95.

Women in the Scientific Search: An American Bibliography 1724-1979. By PATRICIA JOAN SIEGEL and KAY THOMAS FINLEY. Scarecrow Press, 52 Liberty St, PO Box 656, Metuchen, NJ 08840, USA:1985. Pp.399. ISBN 0-8108-1755-1. \$32.50.

Computer-assisted model building

from W.C. Ripka

Developments in computer graphics and distance geometry have provided important tools for comparative model building of protein structures.

THE explosion in the number of available protein sequences derived from gene sequence information has created a wealth of opportunities for molecular modellers. Ideally one would like to use only these amino acid sequences and some set of structure-based rules to arrive at accurate secondary and tertiary structures. Recent efforts using an artificial intelligence rule-based system have shown promise for one structural class, the α - β proteins. Accurate structures were found within families generated by this approach¹. In the general case, however, even secondary structure predictions are often less than 60 per cent accurate. The problem of protein folding, that is, the determination of a protein's tertiary structure from its amino acid sequence, is one of the major unsolved problems of molecular biology.

The problem becomes more manageable, however, if a few assumptions are made. The few hundred protein structures that have been solved by X-ray crystallography show that proteins may be classified in a relatively small number of structural classes. From one of these classes an appropriate homologous member can sometimes be chosen for comparative model building. This assumes that the protein structure to be determined is not too different from the known example, and that changes in side chains and residue insertions and deletions can be considered relatively minor perturbations of the structure as a whole. Such a model can then be used to direct synthetic efforts in drug design and protein engineering.

Model building, particularly of structures as complicated as proteins, has been greatly aided by the increasing number of commercially available computer graphics programs that allow interactive, real-time three-dimensional manipulations of complex molecules² [Some of the available software packages: Mogli (Evans and Sutherland), Insight (Biosym), Biograph (Bio-design), Chemmovie (Chemical Design), Midas (Univ. of Cal., San Francisco), Frodo (Rice Univ.) MMS (Univ. of Cal., San Diego), Syble (Tripos)]. These programs can perform bond rotations, on-line calculations of contact distances, side-chain alterations and residue substitutions. Both solvent-accessible and van Der Waals surfaces can be generated to display the space-filling properties of the molecules. Since these surfaces can be selectively coloured to represent partial atomic charges of the atoms that contribute to each surface element, these

models also give an indication of charge distributions.

In addition to the increase in available software, the rapid fall in computer hardware prices has made these facilities available to both industrial and academic laboratories.

Bases for structure

Comparative model building can offer insight into the relationship between amino acid sequences and three-dimensional structure, but its ultimate promise lies in its ability to elucidate the molecular basis of enzyme-substrate and protein ligand recognition and specificity. Such knowledge would provide the basis for rational and productive drug design. Although model building should be most effective when the protein of interest is closely related to the macromolecule for which detailed structural information is available, successful examples of molecular modelling have ranged from very qualitative models to very precise, comprehensive models.

At one extreme, Ondetti *et al.*³ used primarily chemical and biochemical con-

affinity. These compounds are now in clinical trials as anti-sickling agents.

Comparative model building addresses the cases in between these extremes, by constructing useful models of structures from limited information. For example, using the three-dimensional structure of homologous aspartic proteinases and the primary structure of human renin, Blundell and co-workers^{4,5} have employed interactive computer graphics to propose a tertiary model of human renin. This model may prove of value in the design of specific renin inhibitors.

Greer⁷ has compared the structures of chymotrypsin, trypsin and elastase to allow the construction of 'new' serine proteases based on structurally-conserved and variable regions of these proteins. Using computer graphics, coordinates for the main chain atoms of the model are taken from the conserved region, while side chains are substituted to fit the new protein. Modelling of variable regions is considerably more difficult since these areas involve insertions and deletions for which there is no exact correspondence. In these cases, main chain conformations are required, rather than simple side-chain substitutions. Feldman *et al.*⁸ have suggested some simple techniques using interactive computer graphics to approach these problems.

Distance geometry

A potentially more powerful approach for modelling both side-chain and variable region conformations resides in the methodology of distance geometry⁹, which allows the construction of random conformations that meet selected distance constraints. These constructs can then be evaluated according to energetic considerations and other geometric criteria.

O'Neil and De Grado have used distance geometry methods¹⁰ along with interactive computer graphics to construct a model of calmodulin based on the structurally related intestinal calcium binding (ICB) protein (Fig. 1). In this study, distance geometry was used to construct one of the calcium binding loops by holding the main-chain atoms of the model fixed to the coordinates of the known ICB protein and allowing the side chains to adopt numerous conformations; ligand oxygen atoms were constrained to those positions required in the approximate octahedral geometry of the coordination sphere of calcium. A number of conformers were generated, from which the lowest energy

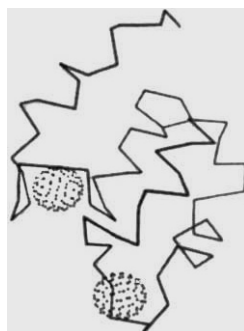


Fig.1 A computer graphics representation of the ICB protein which provided the starting point for calmodulin modelling. The dot surfaces show calcium locations.

siderations to suggest a qualitative relationship between the active sites of carboxypeptidase A, whose structure was known, and angiotensin converting enzyme (ACE), for which no structural information was available. A number of potential enzyme inhibitors of ACE were designed, one of which, captopril, proved to be a potent inhibitor and an effective antihypertensive agent. At the other extreme, Beddell *et al.*⁴ used the X-ray crystal structure of human haemoglobin to design compounds for that molecule. Their modelling suggested an allosteric site at which compounds designed for this site would modulate the protein's oxygen



Fig. 2 O'Neil and De Grado's calmodulin model.

examples (calculated by molecular mechanics) were selected. Of these, only one could be modelled into the proposed calmodulin structure while maintaining reasonable contacts with the rest of the protein (Fig. 2).

Limitations

Although protein models are useful, they may fall short in detailed structural interpretations. The success of comparative modelling depends on correct sequence alignment of the proposed structure with the known example. Even a single mistake can lead to an incorrect result. James and co-workers¹¹ have evaluated the types and sizes of errors in such model building based on a comparison of two proposed models for *Streptomyces griseus* trypsin with the experimentally-determined structure. A second sobering report comes from a study of the structures of bovine and porcine phospholipase A2, in which a single residue change from valine to phenylalanine in an otherwise extended homologous chain, leads to a major reconfiguration of the chain¹². Whether those kinds of changes are predictable remains to be determined. □

W.C. Ripka is at the Biomedical Products Department, Medicinal Chemistry Division, E.I. DuPont de Nemours Co, Wilmington, Delaware 19898, USA.

1. Cohen, F.E., Abarbanel, R.M., Kuntz, I.D. & Fletterick, R.J. *Biochemistry* **22**, 4894-4904 (1983).
2. Langridge, R., Ferrin, T.E., Kuntz, I.D. & Connolly, M.L. *Science* **211**, 661-666 (1981).
3. Ondetti, M.A., Rubin, B. & Cushman, D.W. *Science* **196**, 441-444 (1977).
4. Beddell, C.R., Goodford, P.J., Norringotn, F.E., Wilkinson, S. & Wootton, R. *Brit. J. Pharmac.* **57**, 201-209 (1976).
5. Blundell, T., Sibanda, B.L. & Pearl, L. *Nature* **304**, 273-275 (1983).
6. Sibanda, B.L. *et al.* *FEBS Letters* **174**, 102-111 (1984).
7. Greer, J., *J. molec. Biol.* **153**, 1027-1042 (1981).
8. Feldman, R.J. *et al.* in *Macromolecular Structure and Specificity* (eds Venkataraghavan, B., Feldman, R.J.) 12-43 (New York Academy of Sciences, New York, 1985).
9. Crippen, G.M. in *Chemometric Research Studies Series* Vol. 1 (ed. Baudan, D.) 1-58 (Wiley, New York 1981).
10. O'Neil, K.T. & DeGrado, W.F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4954-4958 (1985).
11. Read, R.J., Brayer, G.D., Jurasek, L. & James, M.N.G. *Biochemistry* **23**, 6570-6575 (1984).
12. Dijkstra, B.W., Weijer, W.J. & Wierenga, R.K. *FEBS Lett.* **164**, 25-27 (1983).

Chemical design takes to the terminals

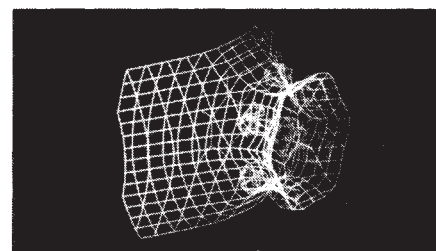
Software packages for chemists turn ordinary microcomputers into tools for molecular modelling, structural analysis and reaction planning.

DEMONSTRATION kits for Molecular Design Ltd's series of PC software provide a low-cost indoctrination into **chemical database programs**. Beginning last month, MDL now offers their ChemBase database manager, designed for IBM PC's in a \$50 (US) demo version that allows the user to experiment with features such as custom table editing, chemical drawing templates, and multi-condition data searching (Reader Service No. 100). ChemText, a chemical wordprocessor package, and the mainframe hook-up ChemTalk/ChemHost, are also available as demo kits. These programs give graphical representations of molecules and reactions in the same format as MDL's mainframe database REACCS (a reaction access system) and MACCS (a molecular access system). CHEMLAB-II, a computational chemistry package, completes MDL's mainframe threesome. The microcomputer demo kits' cost will be deducted from the price of PC software purchased within 30 days of the demo package.

A California company has taken the lead in **three-dimensional molecular modelling** that is both useful and entertaining. Stereographics Corp. has a \$20,000 (US) graphics display system that uses alternating viewpoints and visors with liquid crystal shutters to create a convincing 3-D image (Reader Service No. 101). Stereographics' 3Ddisplay accepts output from any computer and works with all 3-D graphics software

packages. Its imagery provides insight into hidden surfaces with twice the number of fields per second that previous stereoscopic viewers have afforded, reducing irritating flicker effects. A 3-Display system includes a display controller, a 120 Hz RGB monitor and viewing visors.

Chemical Design has a **molecular modelling system** that will run on DEC VAX

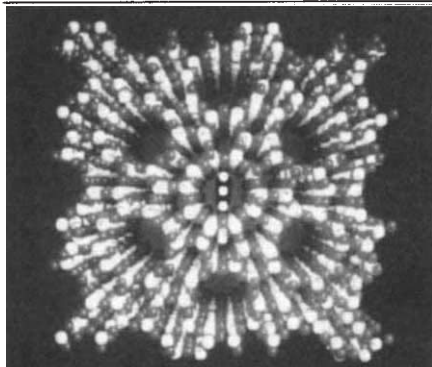


Modelling by Chemical Design. See opposite for other examples.

computers under the VAX/VMS operating system. Chem-X integrates five specialized modules for design, manipulation and analysis (Reader Service No. 102). The system has a user-friendly front end and interfaces to Quantum Chemistry Program Exchange (QCPE) software (see below) as well as MACCS and CHEMLAB-I. Standard supported graphics terminals or PCs with graphics options can run Chemical Design's 4-megabyte program. Chem-X can be incorporated in a combined hardware/software package called MicroGRAF-2,



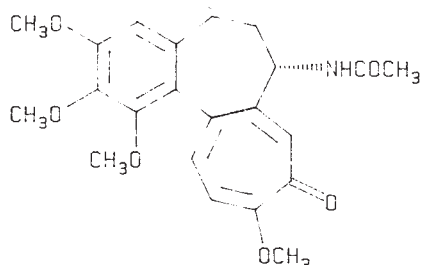
Head-up display in 3-D from Stereographics.



which is priced according to application.

The most affordable **modelling system** for IBM desktops comes from Xiris Corp. in the form of the \$1,000 (US) XICAMM software (Reader Service No. 103). XICAMM can perform many of the tasks of much more expensive systems (albeit more slowly) including molecule comparisons, bond angle determinations, structure rotation and generation of QCPE's MOPAC- or MM2-compatible input data files. An evaluation system with full documentation is also available for users who want to gauge XICAMM's merits for themselves. Xiris provides any or all of the hardware needed to run their system, which is based on the same algorithms used in certain mainframe and super-mini computers.

An unconditional 30-day money-back guarantee accompanies Molecular Pre-

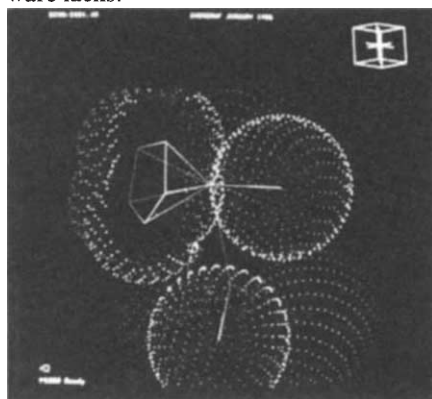


A plot straight from the HP 7470 plotter.

sensation Graphics, the **chemical drawing package** from Hawk Scientific Systems (Reader Service No. 104). Hawk's \$225 (US) package is compatible with the IBM PC and Hewlett-Packard-150, among others, and excels at the speedy generation of quality diagrams for research texts. MPG is not a modelling system, but can supplement one or stand on its own as a useful aid at the write-up stage of research projects. The menu-driven software gives clear drawing instructions, and diagrams can be printed on Epson or HP LaserJet and ThinkJet printers via MPGLASER, a Hawk option that merges text and diagrams.

Chemdata Ltd makes **molecular graphics packages** for Apple II/Ie and BBC model B/Master 128 microcomputers. Their integrated hardware/software unit, including a graphics display processor unit and (for

the Apple series) a hard-wired multiplier card, gives 512 by 512 pixels displays in eight colours (Reader Service No. 105). A structure file editor, model builder, graphics module and library of structure files comprise the software for these units. By merging Chemdata's structure files, two chemical structures can be manipulated together. Chemdata's hardware/software duo costs £890 (UK); in addition, software files with file editors, graphics modules and structure libraries are also available for BBC Master 128 and Apple machines. The Master 128 version, at £75, is priced slightly higher than the £45 Apple series package, but offers several options, such as file merging and model building on a double-buffered four-colour graphics mode, that the less expensive Apple software lacks.



The chemistry department at Indiana University runs a **'library' of chemical analysis software** for IBM and compatible computers called the Quantum Chemistry Program Exchange (QCPE) (Reader Service No. 106). Most of QCPE's offerings are molecular mechanics packages adapted from mainframe programs, which the exchange distributes free of charge to non-profit institutions. Manager Richard Counts conceived the program only a few years ago, since when it has amassed 10 varieties of software for the chemist, including the well-known MOPAC, MM1, MM2 and Galsien 80. Counts plans to expand the program to include a broader range of chemical design packages.

Disks over index cards

The final component of ISI's Sci-Mate version 2.0 software system has now joined the fleet. The editing function for **bibliographic formatting** rounds out Sci-Mate's search and management functions for researching journal literature, building text files and preparing manuscripts (Reader Service No. 107). ISI's package comes in MS-DOS or CP/M 80 formats and cost \$399 (US), with volume discounts available.

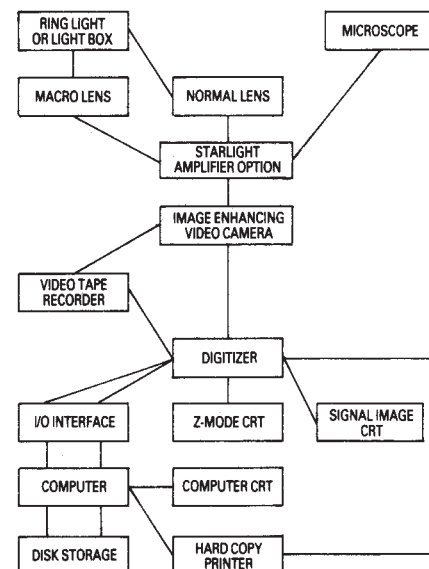
Autoscribe Ltd. distributes a **chemical filing system** for IBM PCs or Hewlett-Packard-150s with a minimum of 256K of

Project Account	
REPT PC	11
PROD PC	12
CLEAR PC	13
FINISH	14
MPG DISP	15
PC	16
PT	17
PAID	18

internal memory. Datalyst entries can store compound names, molecular formulas, chemical formulas, vendor information, laboratory storage location, quantity in stock and details of handling requirements (Reader Service No. 108). The £295 (UK) program will also calculate molecular weights and access chemical structure files from Molecular Presentation Graphics systems (see above). Demonstration disks are available.

Software branches out

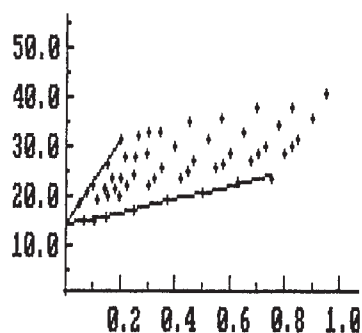
Five new **image analysis programs** from Darwin Instruments, Inc. link Colorado Video's 270A video digitizer with Apple II+ and IIe PCs. The programs were designed for "researchers limited by budget constraints," says Colorado Video (Reader Service No. 109). A scanning densitometer program will perform sum-averaging and multi-scan, baseline addition or subtraction, noise elimination, area integrations and peak identification. A topography program allows Darwin's system to count the total number of objects in a scan field, determine the area and perimeter of each object, calculate the total area of all objects, and delete



Darwin's route to analysed images.

objects based on size class or object number. Darwin also offers an area program for individual object or total area calculation, a program for automating growth curves for plants and sessile animals and a tracking program which can monitor the rate and direction of an object's movement.

In less than 1 minute, software from Wyatt Technology can generate a Zimm plot and **calculate molecular weights** using data from DAWN laser light scattering



Wyatt's Zimm plot.

instruments. Wyatt's package has a menu which allows users to choose different order polynomials for extrapolating the Zimm plot curves (Reader Service No. 110). Instead of being limited to linear extrapolation, the program can handle quadratics, cubics and higher-order expressions. Wyatt's \$1,000 (US) software was developed for IBM PCs, XTs and ATs. Several user inputs, such as the number of concentrations, solvent refractive index and scaling factors for the graphics, must be manually entered for the program to go to work.

Varian Associates put two programs together in a software package that facilitates **automatic data transfer** between atomic absorption spectrometers and a variety of computer systems. With

Varian's \$2,000 (US) data communications package, the user can specify how much information needs to be transferred, and send data on-line or as a summarized report (Reader Service No. 111). An external communications module accepts information and instructions from a remote keyboard or disk file, then transmits results through an IEEE or Centronix port or to remote computers via either or two RS-232C channels. The other half of Varian's package is the public domain program KERMIT, which can transfer files from the spectrometer to micro- or mainframe computers operating with appropriate versions of the program. IBM PC, DEC VAX and Rainbow, HP-110 and HP-150 and WANG PC are among the compatible computers for these data communications.

Microbore tubing up for grabs

Elkay Products' latest catalogue contains \$100 (US) worth of coupons for new products including autoclavable pipette tip racks, 1.5 ml microcentrifuge tubes and 50 ml sterile centrifuge tubes (Reader Service No. 112). In 48 colour pages, *Elkay 86* presents a complete line of plas-



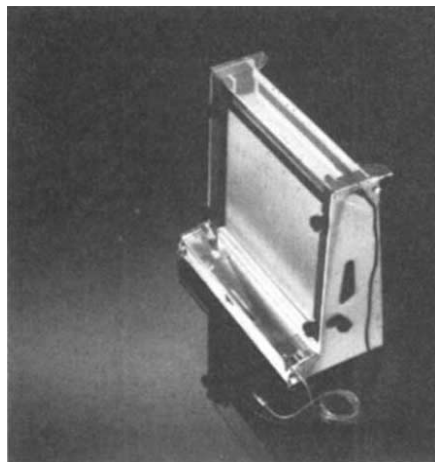
Elkay's latest.

tic disposables for laboratories. General reference material, such as an equipment/consumables compatibility chart, accompanies technical specifications and performance data for Elkay's products.

Elkay is also giving away samples of their extruded microbore tubing (Reader Service No. 113). Sample packages permit inspection of tubing material quality and standards for flow, physical dimensions, durometer and resilience, which Elkay says they carefully control. One package contains six different kinds of tubing for a variety of applications including pump, transmission and bellow forms.

Gel sequels

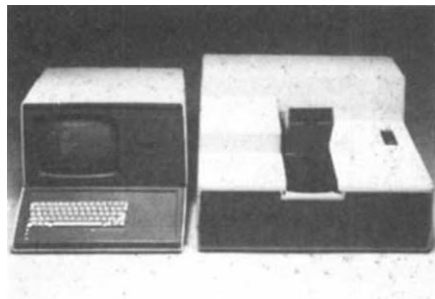
A **sequencing gel electrophoresis system** from Bethesda Research Laboratories promises precision with 48 lanes of read-



Barely \$10 a lane in the BRL S2.

able data for less than \$500 (US). The model S2 apparatus has an aluminium plate for even heat distribution and 0.4 mm side spacers that won't interfere with gel polymerization (Reader Service No. 114). A removable tray allows easy buffer disposal, and the entire unit is constructed of chemical-resistant, unbreakable plastic. BRL's system comes with an accessory package.

Ciba Corning's **scanning densitometer** can analyse up to 64 samples at a time without operator attendance. The model 780 fluorometer/densitometer is pre-programmed for Ciba Corning procedures, but the protocol can be altered for additional tests (Reader Service No. 115). The system's editing function allows the addition or deletion of minima, baseline adjustment, peak deletion or magnification and repositioning of minima. If used in conjunction with Ciba's 16-track agarose gels, the 780 can analyse all 64 samples in less than 12 minutes. An alphanumeric keyboard, 12-inch CRT and fluorescent pre-viewer have been built in to Ciba's unit. All types of electrophoretic media



Sixty-four samples a time in the model 780.

can be accommodated, and the operator can select the units in which results should appear.

Narrowing the scope to **DNA sequenc-**

ADVERTISEMENT

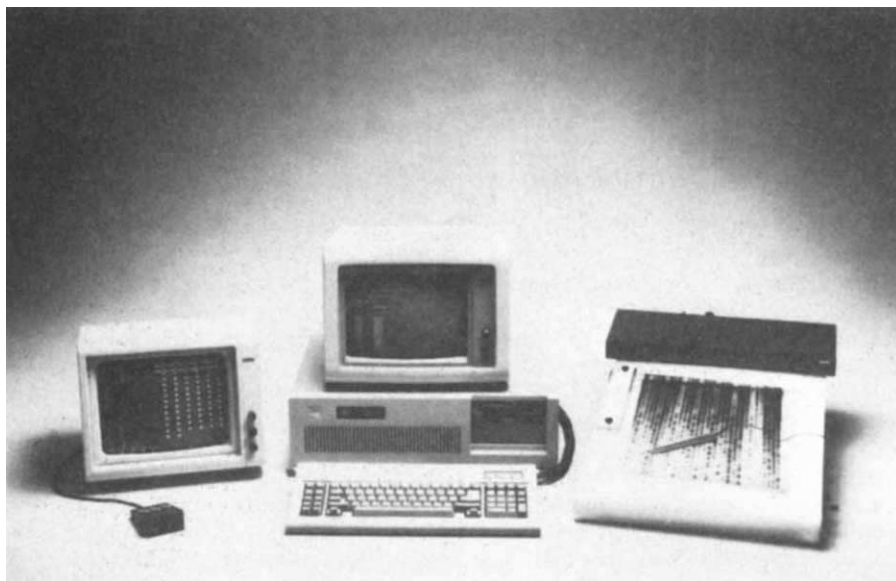
SINGER INSTRUMENTS Xenopus Oocytes injection?



SINGER MKI MICROMANIPULATOR

We also offer a Microsyringe Kit.

Write to:
Singer Instrument Co. Ltd.
Treborough Lodge
Roadwater, Watchet
Somerset TA23 0QL
England. Tel. (0984) 40226
Telex: 337492 Comcab G



ing, Bio-Rad's Gene-Master has been designed for speed and convenience in one particular application (Reader Service No. 116). The system includes an IBM PC AT microcomputer that stores and analyses sequences entered via a digitizing pen. Bio-Rad's model also uses a voice synthesizer to call out bases as the pen records them, so that sequences can be verified without looking away from the autoradiogram. Lane and band curvature doesn't throw the Gene-Master off track, and the menu does not require pen activation which Bio-Rad has deemed distracting. As with most pen-operated readers, the reading process is accelerated when the user can keep an eye on the gel.

Patrons of Shandon's Electrafor products for electrophoresis can now **order supplies by phone** with the Cheshire company's free telephone order service (Reader Service No. 117). A rapid delivery service, combined with the direct-dial



ordering system, helps speed up restocking of the cellulose acetate membranes, buffers and stains in the Elecrafor range. First-time customers will get free samples

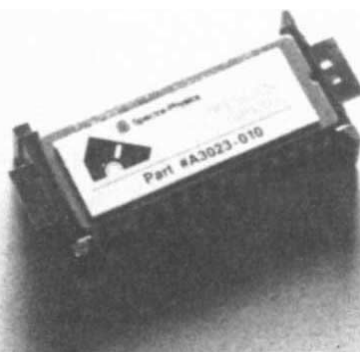
These notes are compiled by Karen Wright from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal.

from Shandon, and all customers get a free box of Electrafor membranes with every 10 purchased. Until the end of May, buffers and stains ordered with membranes will also be discounted 10 per cent.

A New York-based mail-order company has moved into the scientific arena. Transcat Inc. has published its first issue of **Laboratory & Scientific Instruments**, a 274-page catalogue that describes the products of over 200 manufacturers (Reader Service No. 118). Companies featured include Bausch & Lomb, Beckman, Corning, Ohaus, Orion Research, Techné, Thermolyne and Wallace & Tiernan, and products range from microscopes to hot plates. Transcat gives prices, ordering information and a toll-free number to facilitate inquiries.

Tomorrow's program

Spectra-Physics has put a program for **automated stop-flow UV-Vis scanning** on a chip that plugs into their SP4200 computing integrator (Reader Service No. 119). The integrator is attached to a Kratos Spectroflow 783 detector through a current loop serial connection. The program, on the \$1,500 (US) 783 SCAN chip, will perform background subtraction, peak ratioing and time-programmed wave-



Data acquisition by word of mouth

In over 2,500 laboratories across the United States, microcomputers that once turned deaf ears to their surroundings may soon be called on to perk up and listen. Laboratory Technologies, makers of the data acquisition package LABTECH NOTEBOOK, recently agreed to develop a scientific vocabulary for their software commands that would enable the package to respond to voice commands. LT's arrangement with Kurzweil Applied Intelligence will link the Kurzweil Voicesystem (KVS), a 1,000-word speech recognition system, with the IBM PCs that run LABTECH NOTEBOOK (Reader Service No. 120).

LT is hoping that setting up and controlling LABTECH NOTEBOOK will be much easier using simple spoken words. The operator talks into a telephone handset or a lightweight headset microphone, and the NOTEBOOK software does the rest: the PC channels will automatically gather information from attached sensors, instruments or gauges, and pass it on for analysis, control and monitoring. A KVS with an IBM PC AT interface costs \$6,500 (US); the \$895 (US) LABTECH NOTEBOOK package can integrate with Lotus 1-2-3, Symphony and RS/1 for detailed statistical analysis and display. □

length changes for selected liquid chromatography components; scanning parameters can be set via the integrator.

Three new **software utility programs** now complement IBM Instruments' CAP 2 software for their chromatography workstation (Reader Service No. 121). One of the programs, a utilities library, can aid in the analysis and display of chromatography data. Another, called Touch & View, manages screens, keyboards and key pad functions to allow up to seven programmes to run concurrently. GC/9630 Emulation, the third program, was designed to replace the standard VDU control terminal component of IBM's GC/9630 chromatograph: it provides comprehensive instrument control during data acquisition. IBM Instruments sells the programs for \$250, \$400 and \$200 (US) respectively. □

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

UMI Article Clearinghouse

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

Bringing home the technology

from Richard Pearson

New technology is now enabling more people to work from home rather than in the traditional workplace.

EMPLOYMENT and working patterns are constantly changing in response to economic and social and technological developments. In the United Kingdom, the recession of the early 1980s has been seen as a watershed in terms of patterns of working, with employers restructuring their workforces and making greater use of temporary and part-time staff.

Since 1983, the main areas for new employment have been for part-time work, where there has been an increase of more than 300,000 jobs, and self-employment, where the number of people has also increased by over 300,000 to total over 2.5 million. Self-employment has been helped by employers, who are increasing their sub-contracting of work activities such as catering, transport and legal work to concentrate on what they are good at. New technology, in the form of cheaper telecommunications and computing, is also enabling more people to work from home, again showing more opportunities for flexible employment patterns.

In theory, the technology offers the opportunity for vast numbers of office-based



workers and others to carry out their work at home using computer links. Jobs in computer programming, word processing, accounting, statistical analysis and management information processing are all amenable to distance working. In the United States, AT&T estimates that up to half of its executive jobs could be carried out from home by 1990, while in San Francisco it has been estimated that one in five commuters could stay at home to work. But such distance working need know no boundaries. American Airlines moved its data entry operation from Oklahoma to Barbados in the West Indies, where 200 keyboard operators now process tickets, using a satellite link to receive the work from the US data-processing centre. Another US company with a similar operation claims "we can do the work in St

Michaels, Barbados for less than it costs us to pay for the floor space in New York". Yet another company flies its paper work to Singapore, where it is keyed in, and then beams the finished work by satellite to its data-processing centre in Australia.

In the United Kingdom, homeworking is not new; cottage industries in light manufacturing and clothing have been around for generations, although such workers are now a minority of the "homeworkers". Much more common are white-collar workers doing clerical work, or working in occupations such as selling or writing where the home is used as the base. There are estimated to be more than 650,000 homeworkers in the United Kingdom.

There are now a number of initiatives in the United Kingdom aimed at allowing and encouraging traditional office-based professional workers to work from home, becoming the 1980's homemaker or "teleworker". The largest and longest running is the operation of 'F' International, a computer consultancy which employs some 900 people, predominantly women, to work from their homes. The staff, computer programmers, systems analysts and project managers, take on assignments for clients of 'F' International and operate as high-tech cottage workers, working in part from home and in part in the clients' premises. A small core of permanent staff organizes and manages the work, but the 800 or so self-employed workers are able to combine domestic flexibility with a career. They are of course aided in gaining regular work to fit in with their own availability by having skills in short supply.

Another scheme that has attracted a lot of publicity is that started by Rank Xerox, which has cut the number of staff employed in its expensive central London offices. Originally some 20 staff, mainly managers and professionals, were helped to hivel themselves off from the company and to sell their services back to Rank Xerox instead of being employees. As well as being paid special bonuses and provided with training and low-cost equipment, the "networkers" as they were called, were guaranteed two days' work per week at consultancy rates for two years. The company saved significant sums on office rental and services, payment of support staff, canteen facilities, pension, company cars and the like. It also

freed itself of a long-term commitment to employing those staff. The networkers gained a sheltered transition to self-employment and starting their own businesses, and flexibility to work in the way they wanted. Many have been able to achieve much higher levels of output, without for example the interruptions of working in a major office complex, and thus to boost their earnings. Other companies also have hybrid schemes of networkers, freelancers and teleworkers although the technology, other than the telephone, has assumed a less important role than many had first imagined.

There are of course problems. Many people look to the work place to provide social interaction and a structure for their lives, and not all homes and families' lives are conducive to working at home. There is also the uncertainty for the individual in working in a new and uncertain way. For the employer, there are new skills to be acquired in "managing" such people and setting payment systems where it is difficult to measure the output. With a traditional employee, this is easier because the relationship is one of buying the individual's time for, say, forty hours per week.

The productivity gains benefit both sides, and while the individual gains flexibility of working pattern the employer gains flexibility over controlling labour inputs and transferring many of the overheads and "risks" to the individual. The homeworkers gain most when skills are in short supply, and if the pattern is to develop in the future a key issue will be "who pays for the training and development of new skills as they are needed?"

Distance and homeworking will increase in the future as more workers and employers seek part-time and flexible employment relationships. The increasing numbers of women seeking work, and the shift away from manufacturing-based employment to services will also facilitate this trend. The rate of expansion will probably not, however, be as quick as some of the publicity suggests. While the lower costs of technology will aid the process, economic and social factors will be the prime determinants of its development, while the main beneficiaries will be those whose skills are in scarce supply. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.